

Candidato:
Dott. Stefano De Tito

Index

	Pag.
<i>Abstract</i>	6
CHAPTER I: NUCLEIC ACIDS ARCHITECTURE: CANONICAL AND NON CANONICAL STRUCTURES	10
<i>1) General overview of nucleic acids</i>	11
<i>2) Non-canonical DNA structures</i>	15
<i>2.1 Triplex</i>	15
<i>2.2 G-quadruplex</i>	16
<i>References</i>	18
CHAPTER II: DNA G-QUADRUPLEX: FUNDAMENTALS AND BIOLOGICAL FUNCTIONS	19
<i>1) G-Quadruplex structures</i>	20
<i>1.1 Fundamentals of G-quadruplex structures</i>	20
<i>1.2 G-quadruplex groove-widths</i>	23
<i>1.3 G-quadruplex metal ion binding</i>	25
<i>2) Biological Interest of G-quadruplex</i>	27
<i>References</i>	30
CHAPTER III: TARGETING G-QUADRUPLEXES	35
<i>1) G-Quadruplex: an attractive anti-cancer target</i>	36
<i>References</i>	38

2) G-quadruplex groove binders: a story that starts with distamycin A	39
References	41
3) Selective binding of distamycin A derivative to G-quadruplex structure [d(TGGGGT)]₄	42
3.1 Introduction	42
3.2 Results and Discussion	43
3.3 Conclusions	51
3.4 Materials and Methods	52
References	55
4) Structural and conformational requisites in DNA groove binding: another piece o the puzzle	56
4.1 Introduction	56
4.2 Results and Discussion	56
4.3 Conclusions	72
4.4 Materials and Methods	73
References	81
5) Tandem application of virtual screening and NMR experiments in the discoovery of brand new DNA quadruplex groove binders	83
References	89
6) A more detailed picture of the interactions between virtual screening hits and DNA G-quadruplex: NMR molecular modelling and ITC studies	90
6.1 Introduction	90
6.2 Results and Discussion	91

6.3 Conclusions	103
6.4 Materials and Methods	104
References	108
7) Shooting for selective drug-like G-quadruplex binders: evidence for telomeric DNA damage and tumor cell death	109
7.1 Introduction	109
7.2 Results and Discussion	110
7.3 Conclusions	120
7.4 Materials and Methods	121
References	127
 CHAPTER IV: INSIGHT INTO NEW DNA TOPOLOGIES: THE G-TRIPLEX	129
1) The G-triplex DNA	130
1.1 Introduction	130
1.2 Results and Discussion	132
1.3 Conclusions	150
1.4 Materials and Methods	151
References	158
 CHAPTER V: PRELIMINARY STRUCTURE-FUNCTION INVESTIGATIONS OF HsLARP4	160
1) La Protein	161
1.1 La protein: an RNA binding protein	161

1.2 HsLa-RNA interaction	163
References	168
2) La related Proteins	170
2.1 LARPs	170
2.2 Role of LARP4	171
References	175
3) Study of interaction between Human La related Protein 4 (HsLARP4) and poly-A15 RNA	176
3.1 Introduction	176
3.2 Results and Discussion	177
3.3 Conclusions	201
3.4 Materials and Methods	202
References	209

Abstract

Targeting of DNA secondary structures, such as G-quadruplexes, is now considered an appealing opportunity for drug intervention in anticancer therapy. So far, efforts made in the discovery of chemotypes able to target G-quadruplexes mainly succeeded in the identification of a number of polyaromatic compounds featuring end-stacking binding properties. Against this general trend, we were persuaded that the G-quadruplex grooves can recognize molecular entities with better drug-like and selectivity properties.

Therefore, to achieve an enhanced knowledge on the structural and conformational requisites for quadruplex groove recognition, distamycin A, the only compound for which a pure groove binding has been proven, has been chemically modified. Isothermal titration calorimetry (ITC) and NMR techniques have been employed to characterize the interaction between a dicationic derivative of distamycin A (compound **1**) and the [d(TGGGGT)]₄ quadruplex structure. Interestingly, the structural modifications of compound **1** decrease the affinity of the ligand toward the duplex, enhancing the selectivity for the quadruplex structures. Further, structural and thermodynamic studies revealed that the absence of coulombic interactions inferred in compound **3**, results in an unprecedented binding position in which both the groove and the 3' end of the DNA are occupied.

In this scenario, with the aim of finding brand new molecular scaffolds able to interact with the groove of the DNA quadruplex [d(TGGGGT)]₄, we performed a successful structure-based virtual screening (VS) campaign. As a result, six

molecules were found to be somehow groove binders. NMR spectroscopy experiments combined with molecular modelling studies, allow for a more detailed picture of the interaction between each binder and the quadruplex DNA. Noteworthy, isothermal titration calorimetry (ITC) measurements on the above-mentioned compounds revealed that **2**, **4**, and **6** besides their relatively small dimensions bind the DNA quadruplex [d(TGGGGT)]₄ with higher affinity than distamycin A, to the best of our knowledge, the most potent groove binder identified thus far.

Among them, the promising derivative **6** (renamed **1a**) was used as a seed for searching similar entities in several commercially available databases and NMR experiments allowed to identify a small focused library of structural analogues with G-quadruplex binding properties. By a back and forth approach, the structural features responsible for G-quadruplex groove recognition were delineated, while isothermal titration calorimetry (ITC) measurements allowed for the identification of chemotypes featuring a tighter binding than distamycin A. Differently from distamycin A, the best binders were also proved to be G-quadruplex selective over duplex. These results propelled the biological characterization of the new ligands demonstrating their ability to induce selective DNA damage at telomeric level and induction of apoptosis and senescence on tumor cells.

Structural variations from the canonical Watson-Crick double helix have specific roles in many important cellular processes, such as DNA packaging, replication, transcription and recombination. These DNA structures are sequence-directed and constitute an alternative layer of the genetic code. Therefore, revealing new DNA structural motifs provides the molecular bases to elucidate

novel functional mechanisms of cell and the way to interact with them. By combining advanced state-of-the-art computations and experiments, a study conducted also under the supervision of Prof. Bertini (CERM, Sesto Fiorentino) led to the identification of a new DNA structural motif, named “G-triplex”. G-triplex can be formed in guanine-rich regions of the genome and is characterized by the formation of G:G:G triad planes stabilized by an array of Hoogsteen-like hydrogen-bonds. This discovery further expands the structural complexity of the genome highlighting once more the high polymorphism of the DNA polymers. This is the first time that DNA is found to assume this kind of topology and the abundant presence of guanine-rich regions in the genome makes imperative to investigate its biological role in the near future.

In a study conducted in Conte’s lab. (King’s College, London) the interactions between human LARP4 and poly-A15 RNA have been investigated. Human LARP4 (HsLARP4) is a cytoplasmic, polyribosome-associated protein involved in promoting mRNA translation and able to bind to oligo(A) RNA and the poly(A) binding protein (PABP). The RNA recognition is mediated by the La module, the conserved feature in all LARP families. However, human LARP4 is notably distinct from genuine La proteins and other LARPs for the lack of conservation of few amino acids that in human La were shown to be critical for 3’-UUU_{OH} interaction, suggesting a divergent mode of RNA recognition. Here, five highly conserved residues in the La module of HsLARP4 were identified through sequence alignment with HsLa, and three of them (Q126, D139, M160) were mutated to Alanine through point mutation. ITC experiments revealed that the mutation of residues 126, 139 and 160 to Alanine in

the La module of HsLARP4, decreased its binding affinity to poly-A15, hence, indicating that these residues of HsLARP4 play key roles in the interaction with RNA.

CHAPTER I

**NUCLEIC ACIDS ARCHITECTURE:
CANONICAL AND NON-CANONICAL
STRUCTURES**

1) General overview of nucleic acids

Nucleic acids are the biological macromolecules that store and transmit genetic information in living organisms. This information ensures the normal development and functioning of an organism. Nucleic acids comprise of two main classes: ribonucleic acids (RNA) and deoxyribonucleic acids (DNA). The structure of DNA plays an essential role in DNA's ability to store and transmit genetic information. DNA is a polymer of simple units, and the genetic information is encoded in the sequence of these monomeric units, the nucleotides. Nucleotides combine together to form DNA strands, and complementary strands of DNA bind together to form a double stranded helix. The integrity of the genetic information is maintained within this DNA structure. This information is converted via several processes (transcription and translation) into various structural, regulatory and functional components *in vivo*¹. Consequently, understanding the factors that influence the structure of DNA is significant for understanding the processes that compromise this information such as through DNA damage. A nucleotide is a chemical compound that consists of three components: a heterocyclic base, a sugar and one or more phosphate groups. The bases of DNA are divided into two groups: purines [adenine (A) and guanine (G)] and pyrimidines [thymine (T) and cytosine (C)]. In RNA, thymine is replaced by uracil (U). Purines consist of a six-membered and a five-membered nitrogen-containing ring, fused together; pyrimidines have only a six-membered nitrogen-containing ring (Figure 1.1)¹.

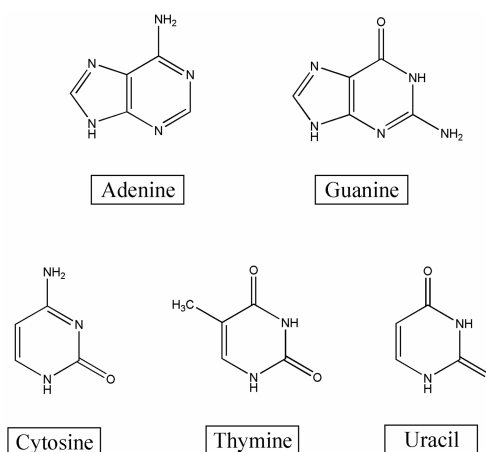


Figure 1.1: Structure of purine and pyrimidine bases.

The sugar component of nucleic acids is a pentose, DNA contains the 2'-deoxyribose, while RNA contains ribose. The molecule comprising the pentose sugar joined to a base is called nucleoside. In a nucleoside, sugar and base are covalently bound by a N-glycosidic bond between the C1' of the sugar and the N9 or N1 of purines or pyrimidines, respectively. Rotations about this glycosidic linkage add structural diversity to the DNA molecule. The two main conformations that are formed about the glycosidic bond are either *syn* or *anti* (Figure 1.2). The *syn* conformation is formed when the C1'-O4' bond is cis to the N9-C4 purine bond, and cis to the N1-C2 of pyrimidine. This occurs when the bulk of the purine base is facing towards the sugar or the C2 carbonyl is on top of the sugar ring. The *anti* conformation is formed when the bulk of the base is rotated away from the sugar. This occurs when the C1'-O4' bond of the sugar is trans to the N9-C4 bond of the purine base. In pyrimidines, the *anti* conformation is formed when the C2 carbonyl faces away from the sugar. This occurs when the C1'-O4' bond is trans to the N1-C2 pyrimidine bond.

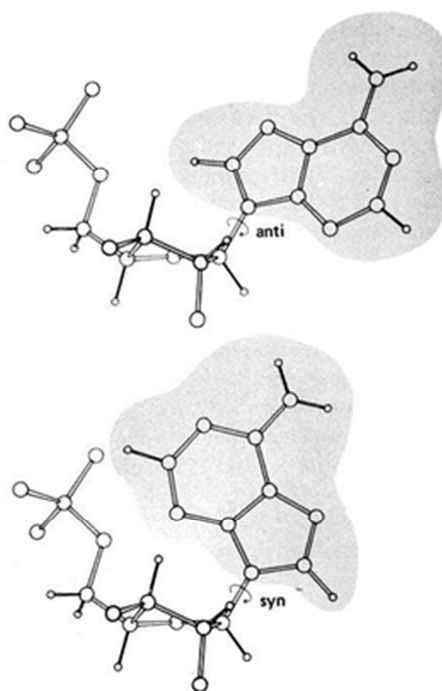


Figure 1.2: Rotation about the glycosidic bond.

A nucleoside is converted into a nucleotide by attachment of a phosphate group to the C5' carbon of the sugar. The nucleotides are connected by phosphodiester bonds to form single-stranded DNA. The phosphodiester bond is formed by the 5' phosphate of one nucleotide attaching to the 3' hydroxyl of the next nucleotide. This attachment lends directionality to DNA. The DNA sequence is usually read from the 5' to 3' direction. The 5' end commonly has a phosphate group attached to it while the 3' end contains just the sugar hydroxyl.

The bases of DNA (or RNA) can form non-covalent bond couplings called base pairs. The most common base pairing pattern found in DNA is the Watson-Crick base pairing. In the canonical Watson-Crick base pairing, adenine forms a base pair with thymine, as does guanine with cytosine in DNA². The base pairs are held

together by hydrogen bonds between covalently bound hydrogen atoms and hydrogen bond acceptor atoms such as carbonyl oxygen atoms or nitrogen atoms. Adenine and thymine (A:T) base pairs are held together by two hydrogen bonds, whereas guanine and cytosine (G:C) base pairs are held by three hydrogen bonds. Although Watson-Crick base pairing is more commonly found in DNA, Hoogsteen base pairing has been observed in several DNA structures such as parallel, triplex and quadruplex DNA structures. Hoogsteen base pairing alters the normal base pairing pattern by base pairing through the N7 position.

Usually, DNA assumes a double stranded helix structure (DNA duplex). The most common DNA duplex conformation is the B-form, which is a right-handed helix². The two complementary strands are anti-parallel, one strand is oriented in the 5'-3' direction while the other strand is oriented in the 3'-5' direction. The bases are within the helix, forming base-pairs that are perpendicular to the axis of the helix. B-DNA has two distinct grooves within its edges due to the orientation of the base pairs: a major groove and a minor groove. The double helix is stabilized and held together by several different forces. Firstly, hydrogen bonding between the bases on the complementary strands stabilizes the DNA duplex. Secondly, hydrophobic base-stacking interactions between the aromatic rings of adjacent base pairs also contribute to stability. Thirdly, water molecules cooperatively bind along the grooves of the DNA adding further stability to the duplex structure. Finally, metal cations surround the negatively charged phosphate groups in the sugar-phosphate backbone adding more stability to the DNA duplex. Over and above B-form, DNA duplex can also assume, under different conditions, a number of other different conformations such as the A-form and Z-form¹ (Figure 1.3).

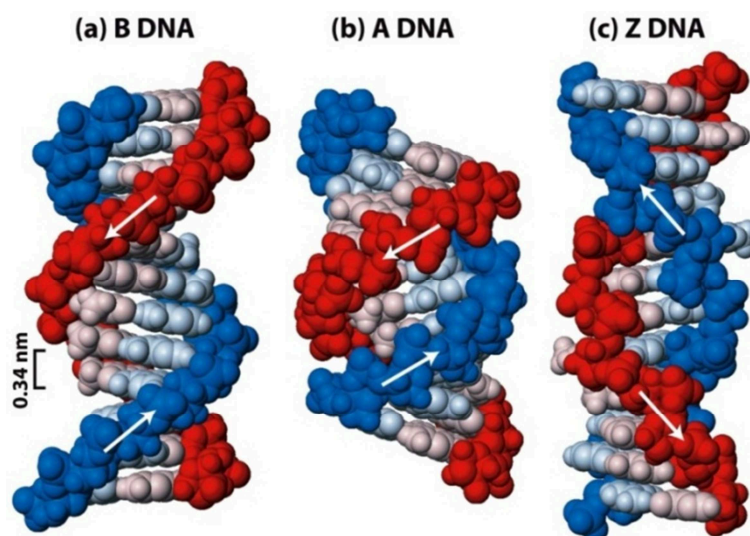


Figure 1.3: Possible forms of DNA duplex: B-form (a), A-form (b), Z-form (c).

2) *Non-canonical DNA structures*

2.1 *Triplex*

The existence of the DNA triple helix structure is known since 1957³, four years after the discovery of the DNA double helix structure², but only recently the triple-helical forms of DNA have received the due attention. The interest in triple helices of DNA is in large part due to the potential biomedical applications for specific controlling gene expression at both transcriptional and replicational levels⁴.

Canonical triple helix formation relies upon hydrogen bonding interaction between a homopyrimidine oligonucleotide and homopurine-homopyrimidine duplex already engaged in Watson-Crick hydrogen bonds. The third oligonucleotide strand (often called TFO, acronym of Triplex Forming Oligonucleotide) occupies the major groove of the double helix forming Hoogsteen hydrogen bonds⁵ with the

purines of the Watson-Crick type base pairs (Figure 1.4). This recognition process is extremely simple, specific and efficient. Specificity is derived from thymine (T) recognition of adenine-thymine base pair (A·T) and protonated cytosine (C^+) recognition of guanine-cytosine base pair (G·C). The efficiency of the single strand binding depends on several factors, such as pH, cation concentration, composition and length of the third strand. NMR⁶, infrared spectroscopy⁷ and fiber diffraction studies⁸ convincingly demonstrated that the original double helix within the triplex adopts a B-like rather than A-like structure as was initially concluded by Arnott and Selsing in 1974⁹.

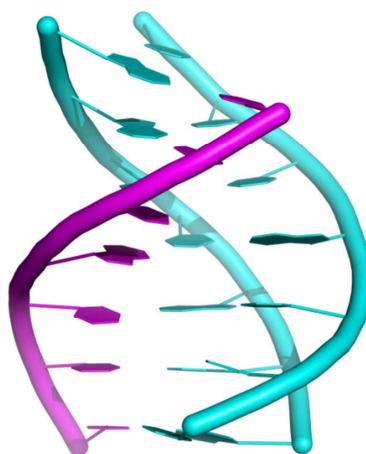


Figure 1.4: Example of G-triplex structure (PDB: 149D).

2.2 G-quadruplex

G-quadruplexes are higher-order DNA and RNA structures formed from G-rich sequences that are stabilized by tetrads of hydrogen-bonded guanine bases (G-quartets) (Figure 1.5)¹⁰. Stacking of G-quartets and coordination of metal cations,

preferentially Na^+ and K^+ , between two adjacent G-quartet planes yield a stable G-quadruplex structure¹¹. Variations in the molecularity, topology, strand orientation, and glycosidic conformation of the G-quadruplex DNA provide a diverse array of structures¹².

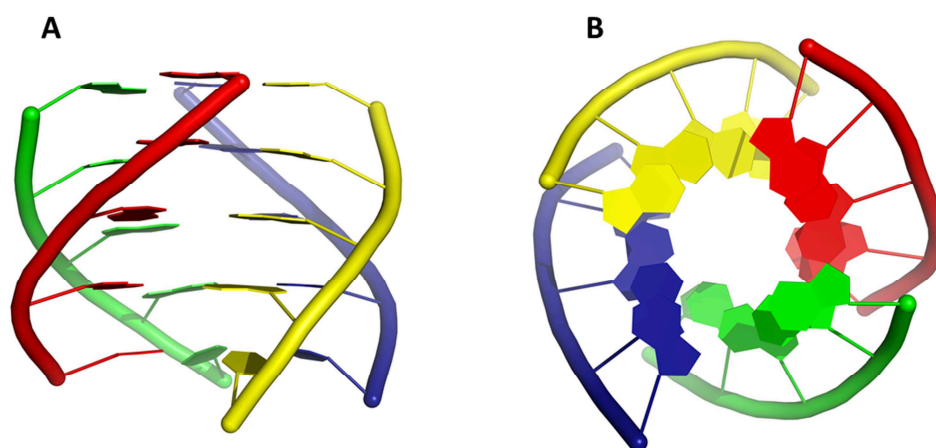


Figure 1.5: (A) Lateral view of a G-quadruplex structure; (B) top view of a G-tetrad.

References

- [1] Voet D., Voet J. G. *Biochemistry*, 2nd Ed. John Wiley & Sons, Inc. **1995**.
- [2] Watson J. D., Crick F. H. *Nature*, **1953**, 171, 737-738.
- [3] Felsenfeld G., Davies, D. R., Rich A. *J. Am. Chem. Soc.*, **1957**, 79, 2023-2024.
- [4] Lyamichev V. I., Mirkin S. N., Frank-Kamenetskii M. D. *J. Biomol. Struct. Dyn.*, **1985**, 3, 667-669.
- [5] Hoogsteen K. *Acta Cryst.*, **1959**, 12, 822-823.
- [6] Macaya R., Wang E., Schultze P., Sklenar V., Feigon J. *J. Mol. Biol.*, **1992**, 225, 755- 773.
- [7] Howard F. B., Miles H. T., Liu K., Frazier J. Raughunathan, G.; Sasisekharan, V. *Biochemistry*, **1992**, 31, 10671-10677.
- [8] Liu K., Sasisekharan V., Miles H. T., Raughunathan G. *Biopolymers*, **1996**, 39, 573-589.
- [9] Arnott S., Selsing E. *J. Mol. Biol.*, **1974**, 88, 509-521.
- [10] Gellert M., Lipsett M. N., Davies D. R. *Proc. Natl. Acad. Sci. USA*, **1962**, 48, 2013-2018.
- [11] Simonsson T. *Biol Chem*, **2001**, 382, 621-628.
- [12] Keniry M. A. *Biopolymers*, **2001**, 56, 123-146.

CHAPTER II

**DNA G-QUADRUPLEX: FUNDAMENTALS AND
BIOLOGICAL FUNCTIONS**

1) G-quadruplex structures

1.1 Fundamentals of G-quadruplex structures

G-quadruplexes can be formed from one, two or four separate G-rich strands of DNA (or RNA) and can display a wide variety of topologies, which are in part a consequence of various possible combinations of strand direction, as well as variations in loop size and sequence¹. They can be defined in general terms as structures formed by a core of at least two stacked G-quartets (or G-tetrads), which are held together by loops that are not involved in the quartets themselves.

The G-quartet consists of a planar arrangement of four guanine bases associated through a cyclic array of Hoogsteen-like hydrogen bonds in which each guanine base both accepts and donates two hydrogen bonds (Figure 1.1)².

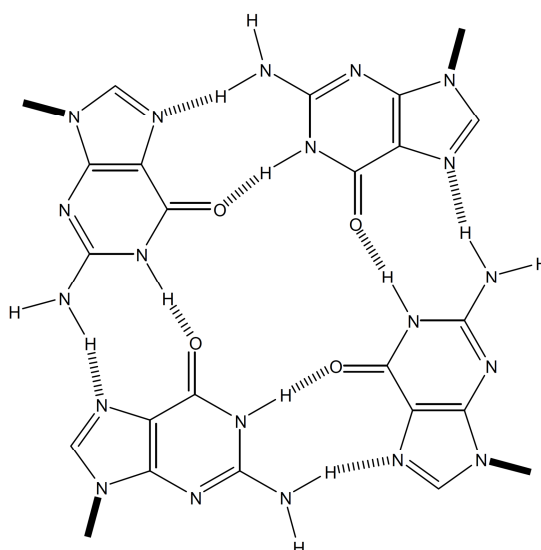


Figure 1.1: Top view of a G-tetrad.

The G-tetrads are not stacked linearly, but adopt a right-handed helix. The combination of the number of stacked G-quartets, the polarity of the strands and the location and length of the loops would be expected to lead to a plurality of G-quadruplex structures, as indeed is found experimentally^{3,4}.

G-quadruplexes can be classified on the basis of: 1) the number of strands (one, two or four); 2) The pattern of strand orientation (parallel, antiparallel); 3) the conformation of guanine glycosidic torsion angles (*anti* or *syn*); 4) the orientation of the loops (lateral, diagonal or propeller) (Figure 1.2)⁵.

Unimolecular quadruplexes are generated by a single oligonucleotide strand whose G-tracts are involved in intramolecular G-tetrad interactions. Potential unimolecular G-quadruplex-forming sequences can be described as follows: $GmXnGmXoGmXpGm$ where m is the number of G residues in each short G-tract, which are usually directly involved in G-tetrad interactions. Xn , Xo and Xp can be any combination of residues, including G, forming the loops. This notation also implies that the G-tracts can be of unequal length, and if one of the short G-tracts is longer than the others, some of the G residues will be located in the loop regions.

Bimolecular and tetramolecular quadruplexes are formed by association of two and four strands, respectively, and, in principle, they can be formed from the association of non-equal sequences. However, almost all bimolecular quadruplexes reported to date are formed by the association of two identical sequences $XnGmXoGmXp$, where n and p may or may not be zero.

Tetramolecular quadruplexes may be formed by four $XnGmXo$ or $GmXnGm$ strands associating together.

X-ray and NMR structures revealed that all the strands of tetramolecular quadruplexes are in a parallel orientation. All parallel quadruplexes have all guanine glycosidic angles in an anti conformation^{6,7}. Quadruplexes are defined antiparallel when at least one of the strands is antiparallel to the others. Antiparallel quadruplexes have both *anti* and *syn* guanine glycosidic torsion angles and their distribution along the strand depends on distinctive topological arrangements, since different topologies have the four strands in differing positions relative to each other⁸⁻¹⁰.

Three types of loops have been observed in quadruplex structures: lateral, diagonal or propeller loops. In general, lateral loops join adjacent G-strands, whereas diagonal loops join opposite Gstrands; in the third type of loops, propeller loops, the loop joins the bottom G-tetrad with the top G-tetrad⁵. In general, unimolecular quadruplexes present all the three types of loop and a variety of antiparallel strand orientations, although parallel strand orientations were also found^{12,13}.

Bimolecular quadruplexes can form different structures depending on even small changes in sequence. For example, the telomeric sequence d(G4T4G4) of the ciliate *Oxytricha nova* forms in solution a symmetric bimolecular structure with parallel/antiparallel strands orientation, diagonal loops formed by opposite strands and alternating *anti/syn* guanine glycosidic angles¹⁴⁻¹⁷. Both sequences d(G3T4G4) and d(G4T4G3) with a missing guanine at the 5' or 3' end with respect to the *Oxytricha nova* sequence, form an asymmetric bimolecular quadruplex. However, the first structure shows one lateral and one diagonal loop¹⁸, whereas the second structure shows two diagonal loops¹⁹.

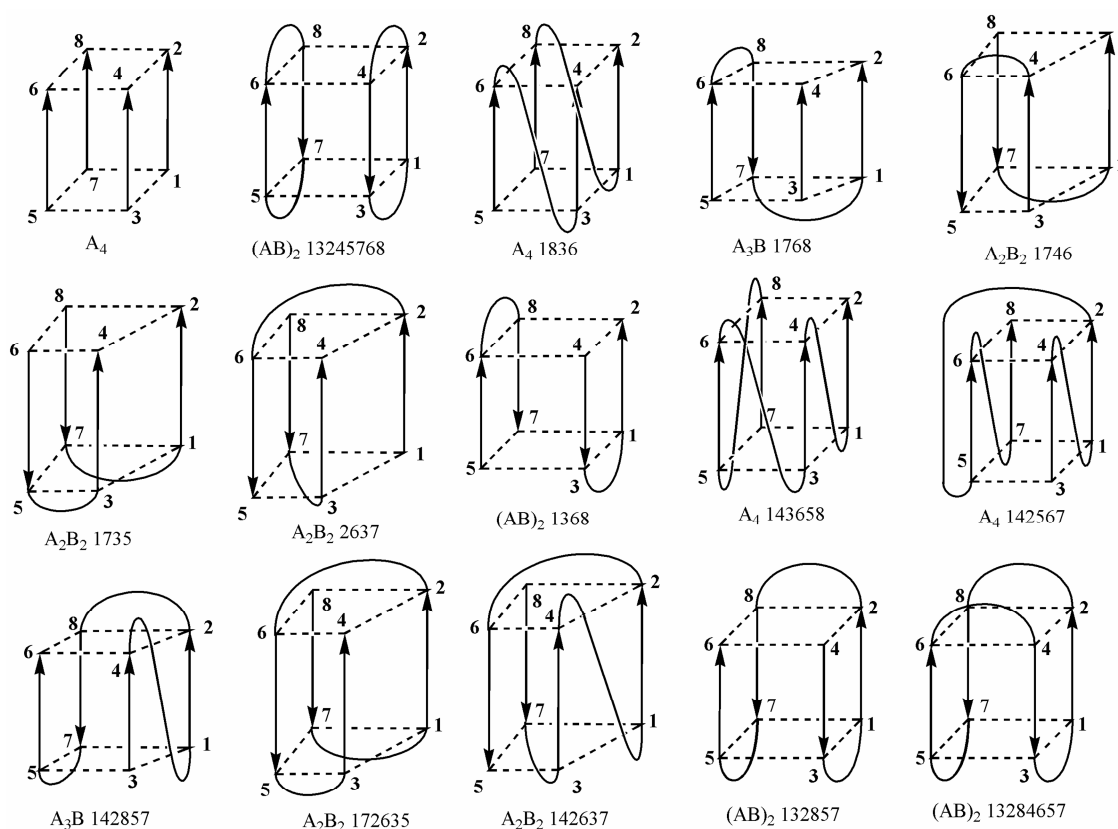


Figure 1.2: Some possible topologies for tetramolecular, bimolecular and unimolecular quadruplexes. The nomenclature indicated is based on a recently proposed topological classification of quadruplexes in which structures are ordered in four main families according to their relative strands orientation (capital letters) and the loops connectivities (Arabic numbers)¹¹.

1.2 G-Quadruplex groove widths

All quadruplex structures have four grooves, defined as the cavities bounded by the phosphodiester backbones. A consequence of variations in the glycosidic torsion angles is the altering of the spacing between the DNA strands of the G-quadruplex, resulting in varying groove-widths along the side of the quadruplex²⁰. In cases where a guanosine in the *syn* conformation donates hydrogen bonds to a neighboring guanosine in the *anti* conformation, the groove formed between the two is extremely narrow, with a phosphate to phosphate distance as small as 7-9

Å²¹. In contrast, when the hydrogen bonding polarity between adjacent *syn*- and *anti*-guanosines is reversed, a very wide groove is formed²¹. An intermediate width groove results when adjacent guanosines adopt the same glycosidic conformation²¹. Indeed, in a parallel four-stranded quadruplex, where all the guanine bases are in the *anti* conformation, the four grooves are approximately the same medium width (Figure 2.2A). For example, in a symmetric bimolecular structure having lateral loops, the G-tetrads are formed from guanosines of alternating *anti-syn-anti-syn* conformation, with each *syn*-guanosine donating hydrogen bonds to an adjacent *anti*-guanosine, and accepting hydrogen bonds from the other adjacent guanosine²². This results in a rectangular G-tetrad core with grooves of alternating wide-narrow-wide-narrow widths (Figure 2.2B). In contrast, a bimolecular quadruplex with diagonal loops, has guanosines that adopt the *syn-syn-anti-anti* conformations, which results in G-tetrads that adopt a parallelogram arrangement and the formation of alternative wide, medium, narrow, medium width grooves between strands (Figure 2.2C).

Groove dimensions are then variable, and depend on overall topology and the nature of the loops. Grooves in quadruplexes with only lateral or diagonal loops are structurally simple, but, in contrast, grooves that incorporate propeller loops have more complex structural features that reflect the insertion of the variable-sequence loops into the grooves.

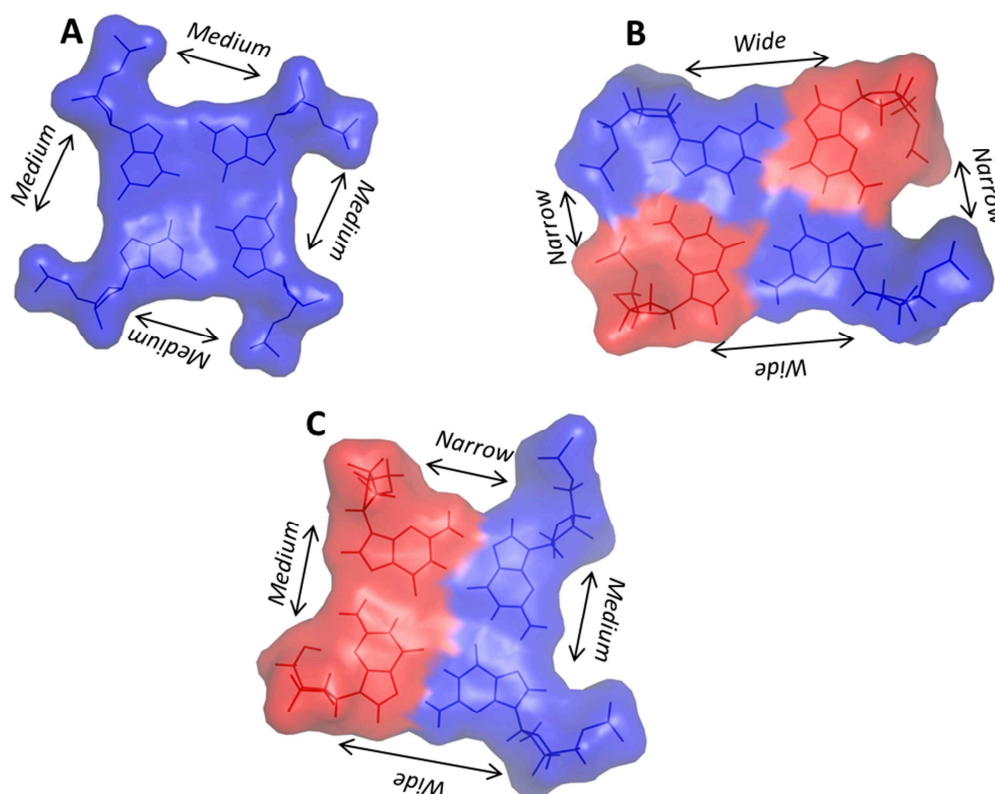


Figure 2.2: Variations in glycosidic torsional angles and their effects on groove width in G-quadruplex structures. All *anti* (A), *anti-syn-anti-syn* (B), and *syn-syn-anti-anti* (C) G-tetrads.

1.3 G-Quadruplex metal ion binding

When one or more G-tetrads are stacked, a cylindrical central cavity is produced²³. This cavity, lined with the guanine O6 carbonyl oxygens, forms a specific binding site for metal ions²⁴. The coordination of potassium²⁵, sodium²⁶, and strontium²⁷ all provide both thermodynamic and kinetic stability to the G-quadruplex structure. Experiments demonstrate that the G-quadruplex is stabilized by the alkali series in the following order: $K^+ > Na^+ > Rb^+ > NH_4^+ > Cs^+ > Li^+$, and for the alkaline earth series in the order: $Sr^{2+} > Ba^{2+} > Ca^{2+} > Mg^{2+}$ indicating that the atomic radii of 1.3 Å of potassium and strontium fit best in the coordination site between

adjacent G-tetrads^{28,29}. The precise location of the cations between the tetrads is dependent on the nature of the ion. Na⁺ ions within the channel have been observed in a range of geometries; in some structures, a Na⁺ ion is in plane with a G-tetrad whereas in others it is between two successive G-tetrads⁷. K⁺ ions are always equidistant between each G-tetrad plane, interacting equally with each of eight carbonyl oxygens in a bipyramidal antiprismatic coordination geometry³⁰.

The selectivity of G-quadruplex DNA for K⁺ versus Na⁺ ion has been studied by Feigon and coworkers by ¹H-NMR, using the bimolecular quadruplex formed by d(G3T4G3) as a model system³¹. In this system, the G-quadruplex binds two K⁺ or Na⁺ ions, corresponding to one metal ion sandwiched between each pair of G-tetrads. Using competition NMR experiments, Feigon and coworkers determined that there is only a modest difference in free energy of 1.7 kcal/mol favouring the binding of K⁺ versus Na⁺ ions. They suggest that this modest free energy difference is a result of the contributions of the relative free energies of hydration, which favors K⁺ binding³¹.

A biologically significant aspect of metal ion binding is related to the observation of different conformation in the presence of different cations. There are a number of well-established examples where the change from Na⁺ to K⁺ induces profound structural alteration, implying high conformational flexibility for these particular quadruplexes. For example, studies on the DNA sequence [d(GGCT4GGGC)]₂ in the presence of either K⁺ or Na⁺ show distinct differences. In the presence of K⁺ ion, two G-tetrads are stacked in an antiparallel form bimolecular quadruplex³².

The smaller Na^+ ion, however, allows for the formation of non-Watson-Crick G:C base pairs in the connecting loop region. These GCGC stabilized tetrads stack on the flanking ends of the G-tetrads with Na^+ ions coordinated between the G-tetrad and its neighboring GCGC tetrad³³.

On the other hand, it is equally clear that some quadruplexes, have very stable and unique topologies. For example, the X-ray structure of the bimolecular quadruplex formed by the *Oxytricha nova* sequence d(G4T4G4) in the presence of K^+ ions¹⁷, shows identical quadruplex topology to the NMR structures in Na^+ -containing solution³⁴⁻³⁶, which is itself identical with the NMR structure in Tl^+ -containing solution³⁴.

2) Biological interest of G-quadruplexes

Recently, there has been growing interest in the study of quadruplexes because of their possible involvement in many biological processes. Telomeric DNA in a variety of eukaryotic organisms is organized in tandem repeats of short G-rich sequences³⁵⁻⁴¹ (Table 2.1) that were shown to form G-quadruplex structures^{12,21,42,43}. The intramolecular G-quadruplex formation observed in the human telomeric sequence d[TTAGGG]₄ can inhibit telomerase activity⁴⁴⁻⁴⁶. As in vitro telomerase induction can transform healthy cells into malignant ones⁴⁷, telomerase inhibition by targeting G-quadruplex leads to telomere shortening and senescence in tumor cells⁴⁸⁻⁵⁷. G-quadruplex structures have been shown to exist in vivo in *Stylonychia lemnae* macronuclei⁵⁸ and recently in human cells⁵⁹.

Group	Organism	Telomeric repeat
Vertebrates	Human, mouse, <i>Xenopus</i>	TTAGGG
Filamentous fungi	<i>Neurospora crassa</i>	TTAGGG
Slime moulds	<i>Physarum</i> , <i>Didymium</i> , <i>Dictyostelium</i>	TTAGGG AG(1-8)
Kinetoplastid protozoa	<i>Trypanosoma</i> , <i>Crithidia</i>	TTAGGG
Ciliate protozoa	<i>Tetrahymena</i> , <i>Glaucoma</i> <i>Paramecium Oxytricha</i> <i>Stylonychia</i> , <i>Euplotes</i>	TTGGGG TTGGG(T/G) TTTTGGGG
Apicomplexan protozoa	<i>Plasmodium</i>	TTAGGG(T/C)
Higher plants	<i>Arabidopsis thaliana</i>	TTTAGGG
Green algae	<i>Chlamydomonas</i>	TTTTAGGG
Insects	<i>Bombyx mori</i>	TTAGG
Roundworms	<i>Ascaris lumbricoides</i>	TTAGGC
Fission yeasts	<i>Schizosaccharomyces pombe</i>	TTAC(A/C)G(1-8)
Budding yeasts	<i>Saccharomyces cerevisiae</i> <i>Candida glabrata</i> <i>Candida albicans</i> <i>Candida tropicalis</i> <i>Candida maltosa</i> <i>Candida guilliermondii</i> <i>Candida pseudotropicalis</i> <i>Kluyveromyces lactis</i>	TGTGGGTGTGGTG (from RNA template) or G(2,3)(TG)(1-6)T (consensus) GGGGTCTGGGTGCTG GGGTACGGATGTCTAACTTCTT GGGTGA[C/A]GGATGTCACGATCATT GGGTACGGATGCAGACTCGCTT GGGTAC GGGTACGGATTGTGATTAGTTATGT GGGTACGGATTGTGATTAGGTATGT

Table 2.1: Sequences and length of Telomeric repeats (Taken from Burge *et al.* NAR 2006)¹⁵

Quadruplex-forming G-rich sequences are also found in a number of cancer-related genes such as c-myc⁶⁰⁻⁶³, c-kit^{64,65}, HIF-1 α ⁶⁶, VEGF⁶⁷, BCL-2^{68,69}, KRAS⁷⁰, Rb^{71,72} and RET protooncogene⁷³. The evidence for the involvement of G-quadruplexes in transcriptional control of the c-myc oncogene has been shown⁷¹. The overexpression of this oncogene is associated with a variety of human malignancies, including breast, colon, cervix, small-cell lung carcinomas, osteosarcomas, glioblastomas, and myeloid leukemias⁷⁴⁻⁷⁷. In addition, c-Myc protein can induce telomerase activity by increasing the transcription rate of hTERT⁷⁸. The nuclease hypersensitivity element (NHE) III1 upstream of the P1 promoter of c-myc controls up to 90% of the transcriptional activation of this gene^{61,79-83}. The element NHE III1 can form transcriptionally active and silenced forms⁸⁴. The guanine-rich segment of this element is part of the silenced form and can adopt a quadruplex structure that is critical for transcriptional silencing^{61,63}.

G-quadruplex interacting agents with specificity for binding to silencer element can be good anticancer agents⁸⁵.

G-quadruplexes are also important because they have been implemented in the design of novel aptamers aimed at binding and inhibiting particular proteins⁸⁶⁻⁸⁸. For example, the sequence d(GGTTGGTGTGGTTGG) (TBA: thrombin binding aptamer) was found to be a potent inhibitor of thrombin in a fibrinogen clotting assay^{89,90}. Furthermore, quadruplex forming oligonucleotides have resulted to be potent inhibitors of the HIV-1 integrase, the enzyme responsible for the insertion of viral DNA into the host genome⁹¹. Recently, several non-telomeric or telomeric G-tail oligonucleotides have been found to exhibit antiproliferative activity against many tumor cells in culture⁹².

References

- [1] Parkinson G. N. In *Quadruplex Nucleic Acids*, S. Neidle; S. Balasubramanian, eds.; RSC Publishing: London, **2006**, pp. 1-30.
- [2] Gellert M., Lipsett M. N., Davies D. R. *Proc. Natl. Acad. Sci. USA*, **1962**, *48*, 2013-2018.
- [3] Keniry M. A. *Biopolymers*, **2001**, *56*, 123-146.
- [4] Williamson J. R. *Annu. Rev. Biophys. Biomol. Struct.*, **1994**, *23*, 703-730.
- [5] Burge S., Parkinson G. N., Hazel P., Todd A. K., Neidle S. *Nucleic Acids Res.*, **2006**, *34*, 5402-5415.
- [6] Aboul-ela F., Murchie A. I., Lilley D. M. *Nature*, **1992**, *360*, 280-282.
- [7] Phillips K., Dauter Z., Murchie A. I., Lilley D. M., Luisi B. J. *Mol. Biol.*, **1997**, *273*, 171-182.
- [8] Strahan G. D., Shafer R. H., Keniry M. A. *Nucleic Acids Res*, **1994**, *22*, 5447-5455.
- [9] Smith F. W., Lau F. W., Feigon J. *Proc Natl Acad Sci U S A*, **1994**, *91*, 10546-10550.
- [10] Smith F. W., Schultze P., Feigon J. *Structure*, **1995**, *3*, 997-1008.
- [11] Esposito V., Galeone A., Mayol L., Oliviero G., Virgilio A., Randazzo A. *Nucleosides Nucleotides Nucleic Acids*, **2007**, *26*, 1155-1159.
- [12] Parkinson G. N., Lee M. P., Neidle S. *Nature*, **2002**, *417*, 876-880.
- [13] Phan A. T., Kuryavyi V., Ma J. B., Faure A., Andreola M. L., Patel, D. J. *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 634-639.
- [14] Smith F. W., Feigon J. *Nature*, **1992**, *356*, 164-168.
- [15] Schultze P., Smith F. W., Feigon J. *Structure*, **1994**, *2*, 221-233.
- [16] Schultze P., Hud N. V., Smith F. W., Feigon J. *Nucleic Acids Res.*, **1999**, *27*, 3018-3028.
- [17] Haider S., Parkinson G. N., Neidle S. *J. Mol. Biol.*, **2002**, *320*, 189-200.
- [18] Crnugelj M., Sket P., Plavec J. *J. Am. Chem. Soc.*, **2003**, *125*, 7866-7871.
- [19] Crnugelj M., Hud N. V., Plavec J. *J. Mol. Biol.*, **2002**, *320*, 911-924.
- [20] Kerwin S. M. *Curr. Pharm. Des.*, **2000**, *6*, 441-478.
- [21] Wang Y., Patel D. J. *J. Mol. Biol.*, **1995**, *251*, 76-94.
- [22] Kang C., Zhang X., Ratliff R., Moyzis R., Rich A. *Nature*, **1992**, *356*, 126-131.

- [23] Sasisekharan V., Zimmerman S., Davies D. R. *J. Mol. Biol.*, **1975**, 92, 171-174.
- [24] Pinnavaia T. J., Marshall C. L., Mettler C. M., Fisk C. L., Miles H. T., Becker E. D. *J. Am. Chem. Soc.*, **1978**, 100, 3625-3627.
- [25] Simonsson T., Sjoback R. *J. Biol. Chem.*, **1999**, 274, 17379-17383.
- [26] Laughlan G., Murchie A. I., Norman D. G., Moore M. H., Moody P. C., Lilley D. M., Luisi B. *Science*, **1994**, 265, 520-524.
- [27] Chen F. M. *Biochemistry*, **1992**, 31, 3769-3776.
- [28] Guschlbauer W., Chantot J. F., Thiele D. *J. Biomol. Struct. Dyn.*, **1990**, 8, 491-511.
- [29] Hardin C. C., Watson T., Corregan M., Bailey C. *Biochemistry*, **1992**, 31, 833-841.
- [30] Strahan G. D., Keniry M. A., Shafer R. H. *Biophys. J.*, **1998**, 75, 968-981. 88
- [31] Hud N. V., Smith F. W., Anet F. A., Feigon J. *Biochemistry*, **1996**, 35, 15383-15390.
- [32] Bouaziz S., Kettani A., Patel D. J. *J. Mol. Biol.*, **1998**, 282, 637.
- [33] Kettani A., Bouaziz S., Gorin A., Zhao A., Jones R. A., Patel D. J. *J. Mol. Biol.*, **1998**, 282, 619.
- [34] Gill M. L., Strobel S. A., Loria J. P. *J. Am. Chem. Soc.*, **2005**, 127, 16723-16732.
- [35] Klobutcher L. A., Swanton M. T., Donini P., Prescott D. M. *Proc. Natl. Acad. Sci. USA* **1981**, 78, 3015-3019.
- [36] Moyzis R. K., Buckingham J. M., Cram L. S., Dani M., Deaven L. L., Jones M. D., Meyne J., Ratliff R. L., Wu J. R. *Proc. Natl. Acad. Sci. USA*, **1988**, 85, 6622-6626.
- [37] Richards E. J., Ausubel F. M. *Cell*, **1988**, 53, 127-136.
- [38] Greider C. W., Blackburn E. H. *Nature*, **1989**, 337, 331-337.
- [39] Kipling D., Cooke H. J. *Nature*, **1990**, 347, 400-402.
- [40] McEachern M. J., Hicks J. B. *Mol. Cell. Biol.*, **1993**, 13, 551-560.
- [41] Prowse K. R., Greider C. W. *Proc. Natl. Acad. Sci. USA*, **1995**, 92, 4818-4822.
- [42] Wang Y., Patel D. J. *Structure*, **1993**, 1, 263-282.
- [43] Wang Y., Patel D. J. *Structure*, **1994**, 2, 1141-1156.
- [44] Mokbel K., Parris C. N., Ghilchik M., Williams G., Newbold R. F. *Am. J. Surg.*, **1999**, 178, 69-72.
- [45] Zhu J., Wang H., Bishop J. M., Blackburn E. H. *Proc. Natl. Acad. Sci. USA*, **1999**, 96, 3723-3728.
- [46] Mokbel K. *Curr. Med. Res. Opin.*, **2003**, 19, 470-472.

- [47] Hahn W. C., Counter C. M., Lundberg A. S., Beijersbergen R. L., Brooks M. W., Weinberg R. A. *Nature*, **1999**, *400*, 464-468.
- [48] Sun D., Thompson B., Cathers B. E., Salazar M., Kerwin S. M., Trent J. O., Jenkins T. C., Neidle S., Hurley L. H. *J. Med. Chem.*, **1997**, *40*, 2113-2116.
- [49] Bryan T. M., Cech T. R. *Curr. Opin. Cell Biol.*, **1999**, *11*, 318-324.
- [50] de Lange T., Jacks T. *Cell*, **1999**, *98*, 273-275.
- [51] Perry P. J., Read M. A., Davies R. T., Gowan S. M., Reszka A. P., Wood A. A., Kelland L. R., Neidle S. *J. Med. Chem.*, **1999**, *42*, 2679-2684.
- [52] Gowan S. M., Heald R., Stevens M. F., Kelland L. R. *Mol. Pharmacol.*, **2001**, *60*, 981-988.
- [53] Mergny J.-L., Lacroix L., Teulade-Fichou M.-P., Hounsou C., Guittat L., Hoarau M., Arimondo P., Vigneron J.-P., Lehn J.-M., Riou J.-F., Garestier T., Helene C. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 3062-3067.
- [54] Mergny J. L., Riou J. F., Mailliet P., Teulade-Fichou M. P., Gilson E. *Nucleic Acids Res.*, **2002**, *30*, 839-865.
- [55] Rezler E. M., Bearss D. J., Hurley L. H. *Curr. Opin. Pharmacol.*, **2002**, *2*, 415-423.
- [56] Riou J. F., Guittat L., Mailliet P., Laoui A., Renou E., Petitgenet O., Megnin-Chanet F., Helene C., Mergny J. L. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 2672-2677.
- [57] Rezler E. M., Bearss D. J., Hurley L. H. *Annu. Rev. Pharmacol. Toxicol.*, **2003**, *43*, 359-379.
- [58] Schaffitzel C., Berger I., Postberg J., Hanes J., Lipps H. J., Pluckthun A. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 8572-8577.
- [59] Chang C. C., Chu J. F., Kao F. J., Chiu Y. C., Lou P. J., Chen H. C., Chang T. C. *Anal Chem*, **2006**, *78*, 2810-2815.
- [60] Simonsson T., Pecinka P., Kubista M. *Nucleic Acids Res.*, **1998**, *26*, 1167-1172.
- [61] Siddiqui-Jain A., Grand C. L., Bearss D. J., Hurley L. H. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 11593-11598.
- [62] Phan A. T., Modi Y. S., Patel D. J. *J. Am. Chem. Soc.*, **2004**, *126*, 8710-8716.
- [63] Ambrus A., Chen D., Dai J., Jones R. A., Yang D. *Biochemistry*, **2005**, *44*, 2048-2058. 89
- [64] Rankin S., Reszka A. P., Huppert J., Zloh M., Parkinson G. N., Todd A. K., Ladame S., Balasubramanian S., Neidle S. *J. Am. Chem. Soc.*, **2005**, *127*, 10584-10589.

- [65] Fernando H., Reszka A. P., Huppert J., Ladame S., Rankin S., Venkitaraman A. R., Neidle S., Balasubramanian S. *Biochemistry*, **2006**, *45*, 7854-7860.
- [66] De Armond R., Wood S., Sun D., Hurley L. H., Ebbinghaus S. W. *Biochemistry*, **2005**, *44*, 16341-16350.
- [67] Sun D., Guo K., Rusche J. J., Hurley L. H. *Nucleic Acids Res.*, **2005**, *33*, 6070-6080.
- [68] Dai J., Chen D., Jones R. A., Hurley L. H., Yang D. *Nucleic Acids Res.*, **2006**, *34*, 5133-5144.
- [69] Dai J., Dexheimer T. S., Chen D., Carver M., Ambrus A., Jones R. A., Yang D. *J. Am. Chem. Soc.*, **2006**, *128*, 1096-1098.
- [70] Cogoi S., Xodo L. E. *Nucleic Acids Res.*, **2006**, *34*, 2536-2549.
- [71] Murchie A. I., Lilley D. M. *Nucleic Acids Res.*, **1992**, *20*, 49-53.
- [72] Xu Y., Sugiyama H. *Nucleic Acids Res.*, **2006**, *34*, 949-954.
- [73] Sun D., Pourpak A., Beetz K., Hurley L. H. *Clin. Cancer Res.*, **2003**, *9*, A218.
- [74] Spencer C. A., Groudine M. *Adv. Cancer Res.*, **1991**, *56*, 1-48.
- [75] Marcu K. B., Bossone S. A., Patel A. J. *Annu. Rev. Biochem.*, **1992**, *61*, 809-860.
- [76] Facchini L. M., Penn L. Z. *FASEB J.*, **1998**, *12*, 633-651.
- [77] Pelengaris S., Rudolph B., Littlewood T. *Curr. Opin. Genet. Dev.*, **2000**, *10*, 100-105.
- [78] Wang J., Xie L. Y., Allan S., Beach D., Hannon G. J. *Genes Dev.*, **1998**, *12*, 1769-1774.
- [79] Bossone S. A., Asselin C., Patel A. J., Marcu K. B. *Proc. Natl. Acad. Sci. USA*, **1992**, *89*, 7452-7456.
- [80] Cooney M., Czernuszewicz G., Postel E. H., Flint S. J., Hogan M. E. *Science*, **1988**, *241*, 456-459.
- [81] Postel E. H., Berberich S. J., Rooney J. W., Kaetzel D. M. *J. Bioenerg. Biomembr.*, **2000**, *32*, 277-284.
- [82] Sakatsume O., Tsutsui H., Wang Y., Gao H., Tang X., Yamauchi T., Murata T., Itakura K., Yokoyama K. *J. Biol. Chem.*, **1996**, *271*, 31322-31333.
- [83] Tomonaga T., Levens D. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*, 5830-5835.
- [84] Collins I., Weber A., Levens D. *Mol. Cell. Biol.*, **2001**, *21*, 8437-8451.

- [85] Seenisamy J., Bashyam S., Gokhale V., Vankayalapati H., Sun D., Siddiqui-Jain A., Streiner N., Shin-Ya K., White E., Wilson W. D., Hurley L. H. *J. Am. Chem. Soc.*, **2005**, *127*, 2944-2959.
- [86] Lin Y., Padmapriya A., Morden K. M., Jayasena S. D. *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 11044-11048.
- [87] Lin Y., Jayasena S. D. *J. Mol. Biol.*, **1997**, *271*, 100-111.
- [88] Wen J. D., Gray D. M. *Biochemistry*, **2002**, *41*, 11438-11448.
- [89] Griffin L. C., Tidmarsh G. F., Bock L. C., Toole J. J., Leung L. L. *Blood*, **1993**, *81*, 3271-3276.
- [90] Li W. X., Kaplan A. V., Grant G. W., Toole J. J., Leung L. L. *Blood*, **1994**, *83*, 677-682.
- [91] Jing N., Hogan M. E. *J. Biol. Chem.*, **1998**, *273*, 34992-34999.
- [92] Qi H., Lin C. P., Fu X., Wood L. M., Liu A. A., Tsai Y. C., Chen Y., Barbieri C. M., Pilch D. S., Liu L. F. *Cancer Res.*, **2006**, *66*, 11808-11816.

CHAPTER III

TARGETING G-QUADRUPLEXES

1) G-quadruplexes: an attractive anti-cancer target

Molecular recognition is a very important phenomenon in biology and chemistry, and the molecular basis of interaction between small ligands and biological macromolecules is the subject of numerous investigations aimed at the rational design of molecules with specific biological activities. In particular, during the last years, a great number of manuscripts have appeared in the literature dealing with the development of low molecular weight agents specifically interacting with DNA¹. Since now, most of the investigations were focused on targeting duplex DNA. Only recently, upon the identification of G-quadruplex motifs as biologically crucial structures, a certain number of studies about ligand/DNA-quadruplex interactions have been reported².

Sequences with propensity to form G-quadruplexes have been identified in biologically significant genomic regions such as telomeres or oncogene promoter regions^{3,4}, which have emerged as potential targets for anticancer drug development.

The biological importance of telomeric G-quadruplex structures arises from the evidence that high telomerase activity (not present in somatic cells) has been implicated in about 85% of tumours⁵. The telomerase elongates the G-rich strand of telomeric DNA, leading the cancer cells to infinite lifetime. For that reason, the inhibition of telomerase has become an interesting strategy for the anticancer therapy⁶. Since the formation of G-quadruplexes by telomeric DNA inhibits the activity of telomerase, small molecules that stabilize the G-quadruplex structures could potentially be effective chemotherapeutic agents⁷. In this scenario, the

identification of new ligands that are specific for G-quadruplex structures is emerging as a promising approach to develop new anticancer drugs. Despite the fact that the structures of G-quadruplexes differ considerably from the double helix, the design of selective quadruplex ligands is very difficult, because the structure of G-quadruplexes varies in several different ways, including number and orientation of strands, grooves width, and loops topology⁸. Nevertheless, a number of G-quadruplex binding agents has been proposed so far and some of these have been demonstrated to be effective telomerase inhibitors⁹.

References

- [1] Nelson S. M., Ferguson L. R, Denny W. A. *Mutat Res.* **2007**, 623, 24-40.
- [2] Ou T. M., Lu Y. J., Tan J. H., Huang Z. S., Wong K. Y., Gu L. Q. *ChemMedChem.* **2008**, 3, 690-713.
- [3] J. L. Huppert and S. Balasubramanian. *Nucleic Acids Res.* **2005**, 33, 2908-2916.
- [4] J. E. Johnson, J. S. Smith, M. L. Kozak and F. B. Johnson. *Biochimie.* **2008**, 90, 1250-1263.
- [5] B. Pagano and C. Giancola. *Curr. Cancer Drug Targets.* **2007**, 7, 520-540.
- [6] J. F. Riou. *Curr. Med. Chem. Anticancer Agents.* **2004**, 4, 439-443.
- [7] S. Neidle. *Curr. Opin. Struct. Biol.* **2009**, 19, 239-250.
- [8] Burge S., Parkinson G. N., Hazel P., Todd A. K., Neidle S. *Nucleic Acids Res.*, **2006**, 34, 5402-5415.
- [9] Martino L., Virno A., Pagano B., *et al.* *J. Am. Chem. Soc.* **2007**, 129, 16048-16056.

2) G-quadruplex groove binders: a story that starts with distamycin A

So far, most of the reported G-quadruplex binding agents bind to DNA by interacting with the wide π -surface of the G-tetrads at the edges of the quadruplex. Groove-binding mode to quadruplex structure was first proposed by our research group for the distamycin A¹, and, up to now, this is the only molecular entity for which a pure groove binding mode has been experimentally proven.

The interest in the groove binding recognition is mainly fueled by the chemical and conformational differences existing between quadruplex and duplex grooves. Unfortunately, the understanding of the chemical, structural and conformational features responsible for G-quadruplex groove binding is still in its infancy. In this scenario, discovery of new distamycin A analogues could greatly contribute to linking structural features and groove binding properties.

Analysis of the distamycin A/[d(TGGGGT)]₄ interaction patterns revealed that this ligand binds, as antiparallel dimer, the quadruplex groove by adopting a crescent shape and by establishing four H-bonds with the G bases¹ (Figure 2.1). Moreover, strong coulombic interactions are established between positively charged amidinium moiety of the ligand and the backbone phosphate groups of the quadruplex.

In a previous study, we investigated the importance of the crescent shape extension by varying the pyrrole units number in distamycin A². Experiments reveal that the presence of one additional pyrrole unit affects the binding of the molecule and varies its stoichiometry, whereas, the addition of two pyrrole units lead to a total loss of affinity of the new derivative for [d(TGGGGT)]₄.

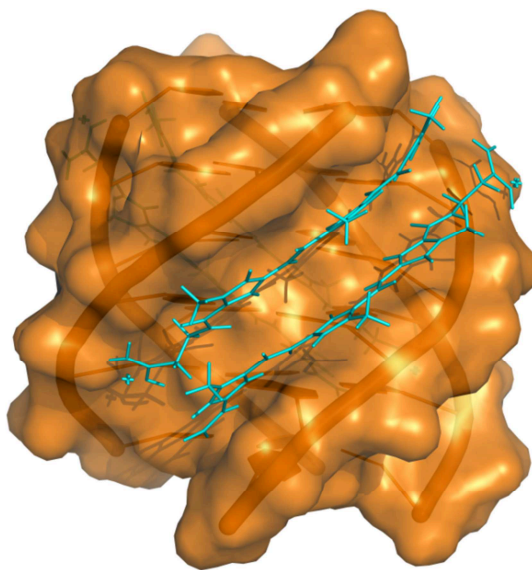


Figure 2.1: 3D structure of the complex 4:1 distamycin A (cyan)/ [d(TGGGGT)]₄ (orange).

In section 3 and 4 of the present dissertation, the description of new chemical modifications on the structure of distamycin A is reported. Particularly NMR analysis, ITC experiments and Molecular modeling studies were used to investigate the interaction of two new distamycin A analogues with the quadruplex structure formed by [d(TGGGGT)]₄.

References

- [1] Martino L., Virno A., Pagano B., *et al. J. Am. Chem. Soc.* **2007**, 129, 16048-16056.
- [2] Pagano B., Virno A.; Mattia C. A., Mayol L., Randazzo A., Giancola C. *Biochimie* **2008**, 90, 1224-1232

3) Selective binding of distamycin A derivative to G-quadruplex structure [d(TGGGGT)]₄

3.1 Introduction

In this work, we report a calorimetric and NMR study of the interaction between the [d(TGGGGT)]₄ quadruplex and a new distamycin A derivative (compound **1**, Figure 3.1), where the initial formamide group is replaced by a charged N-formimidoil moiety.

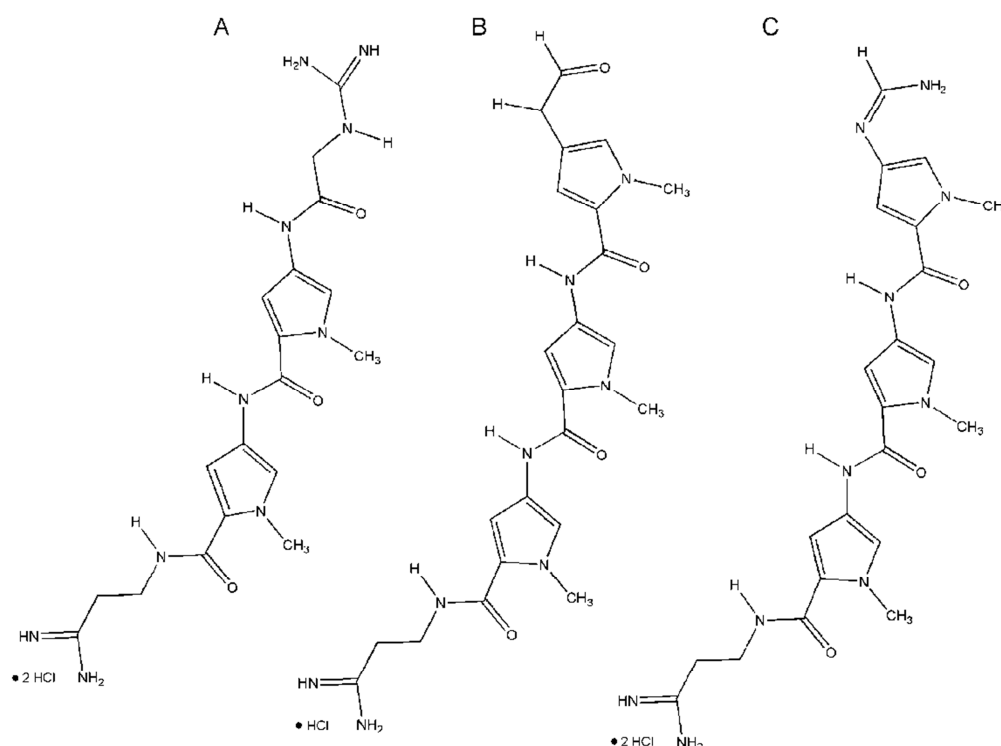


Figure 3.1: Chemical structures of netropsin (A), distamycin A (B) and compound **1** (C).

In addition, since compound **1** can also be considered as an analogue of netropsin because it presents one pyrrole unit more than netropsin (three instead of two), but two cationic ends like it (even if different groups), we compare the

binding of compound **1** to [d(TGGGGT)]₄ with the binding of netropsin to the same target. Finally, to investigate the selectivity of compound **1** for the G-quadruplex relative to duplex, we also performed a study of the interaction between the drug and the selfcomplementary DNA duplex d(CGCGAATTCGCG)₂. This symmetric dodecamer was chosen because (a) it contains the central AATT core, which is considered the specific binding site for distamycin A and netropsin; (b) the interaction with netropsin is well characterized in literature, from both structural and thermodynamic point of view¹⁻⁴.

3.2 Results and Discussion

Distamycin A and netropsin have been recognized for decades as non intercalative DNA binding ligands that show specificity for the minor groove of dA • dT base pairs^{5,6}. The binding of the drugs to duplex DNA involves an electrostatic component from the cationic ends, hydrogen bonds from the amide NH groups, and van der Waals interactions with the wall of the groove. Some years ago, NMR studies indicated that, depending on DNA sequence, some binding sites can accommodate two distamycin A molecules side-by-side in an antiparallel orientation⁷. In this 2:1 complex each ligand molecule preserves all the molecular recognition elements of groove binders. In contrast to distamycin A, the dication netropsin binds only as a single molecule per binding site, suggesting that the side-by-side arrangement of two molecules is inhibited by charge repulsions. Both drugs have been shown little or no affinity for single-stranded DNA or RNA or for double-stranded RNA or DNA RNA hybrids as well as they do not bind to the A helix or the

left-handed Z-DNA⁶⁻⁸. Surprisingly, distamycin A has recently been shown to interact also with four-stranded parallel DNA quadruplexes^{9,10}. Particularly, we have proved, by using NMR and ITC methodology, that distamycin A is able to interact with the quadruplex [d(TGGGGT)]₄. We showed that four ligand molecules bind as antiparallel dimers to the quadruplex in two opposite grooves, establishing hydrogen bonds with the guanine bases and strong coulombic interactions between the positively charged amidinium moiety of the ligand and the backbone phosphate groups of the quadruplex¹⁰. On the other hand, netropsin turned out to possess a lower affinity (NMR data) towards the quadruplex [d(TGGGGT)]₄¹¹, even if, till now, this has never been confirmed by ITC. Interestingly, it seems that netropsin is not able to bind the quadruplex in dimeric form, most probably due to the doubly charged nature of molecule that prevents a side-by-side arrangement into the grooves.

NMR experiments. In order to evaluate the binding properties of derivative **1**, and to perform a direct comparison with the binding behaviour of distamycin A and netropsin, [d(TGGGGT)]₄ has been titrated with **1** at the same experimental conditions (buffer, temperature, DNA concentration) used for distamycin A¹¹. As far as distamycin A is concerned, below 2:1 ligand : quadruplex stoichiometry, the addition of distamycin A to [d(TGGGGT)]₄ caused gradual changing in chemical shift of the signal of the quadruplex, whereas further addition of drug caused the appearance of a new set of proton signals, whose intensities rose by increasing the amount of drug with the concomitant falling off of the original signals which completely disappeared at a ratio of 4:1 drug-DNA. Differently, the NMR titration

profile of **1** (Figure 3.2) turned out to be very different from that observed in the case of distamycin A. The titration was virtually completed at ligand : DNA ratio of 2:1. The four strands resulted to be magnetically equivalent throughout the titration, since no splitting of resonances was observed at any stage. Furthermore, a single set of signals was present for derivative **1** protons throughout the whole NMR titration, which only grew in intensity and did not show any significant change in chemical shift values by increasing ligand concentration. This result clearly suggests that derivative **1** binds the quadruplex in a fast process on the NMR time scale, very similarly to the NMR titration profile observed for netropsin¹¹.

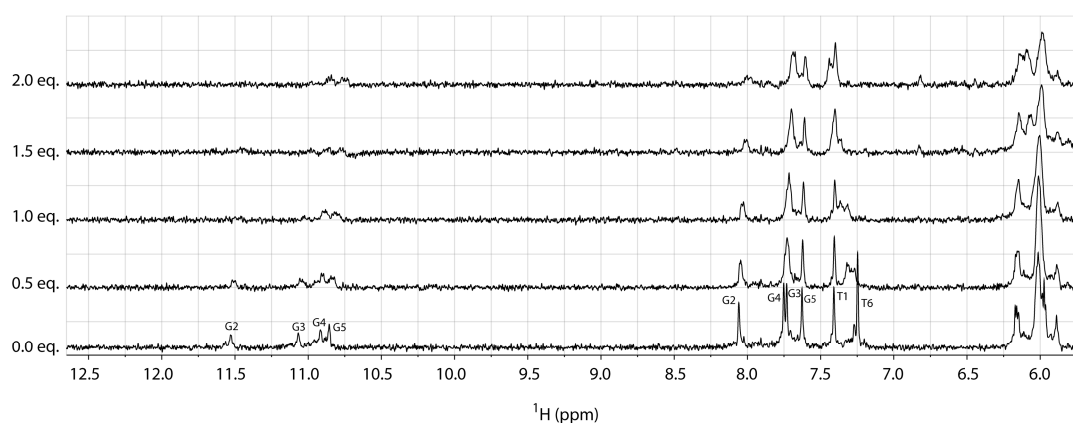


Figure 3.2: NMR titration of [d(TGGGGT)]₄ with derivative **1** (700 MHz, T = 25 °C). 2 mM (8 mM single strand concentration), in 0.2 ml (H₂O/D₂O 9:1) buffer solution having 10 mM KH₂PO₄, 70 mM KCl, 0.2 mM EDTA, pH 7.0. Equivalents of the drug are reported on the left of each spectrum.

In order to preliminarily evaluate the binding site of derivative **1**, a comparison of resonances of some protons of the uncomplexed DNA and the complexed one has been done. In particular, we report the $\Delta\delta$ values (chemical shifts of the complex minus free DNA) of aromatic, methyl, and imino protons in Figure 3.3. Interestingly, all analyzed DNA resonances shifted. Nevertheless, the signal of the protons of T1 residue shifted the least, whereas the ones of residue T6

the more. In any case, a general shift of the aromatic and imino signals could be observed also for the G2, G3, G4, and G5. This means that, basically, derivative **1** is able to recognize most of the molecule surface.

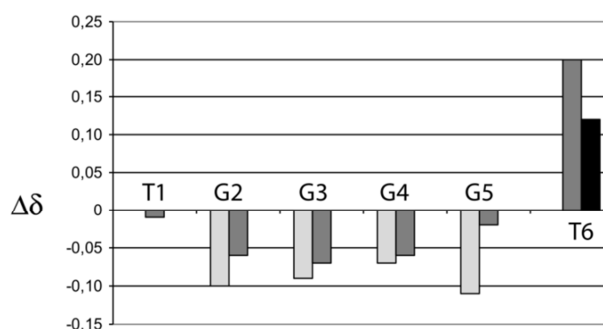


Figure 3.3: Difference in chemical shifts ($\Delta\delta$) of DNA upon binding of **1** (complex **1**:DNA minus DNA alone) with $[d(TGGGGT)]_4$. $\Delta\delta$ values are reported for imino (light gray), aromatic (dark gray) and methyl (black) protons.

We also studied the interaction of compound **1** with the DNA duplex $d(CGCGAATTCGCG)_2$ containing the central AATT core, which is considered the specific binding site for distamycin A and netropsin, to evaluate a possible selectivity of this analogue. The ^1H -NMR spectrum (700 MHz, $T = 25^\circ\text{C}$) of $d(CGCGAATTCGCG)_2$ turned out to be consistent with the assignment already published by Hare *et al.* (taking in due account the different temperature at which it has been assigned)¹². The NMR titration of $d(CGCGAATTCGCG)_2$ with **1** is reported in Figure 3.4. In this case, most of the DNA resonances turned out to be not affected by the addition of ligand, but some resonances of residues C1, A5, A6, T7, T8, C9. In particular, at 2:1 (drug/DNA) molar ratio the aromatic protons of residue C1, A5, A6, T7, T8, C9 residues underwent to a shift of only 0.03, 0.01, 0.01, 0.02, 0.03, and 0.01 ppm, respectively, indicating a very poor affinity towards the dodecamer.

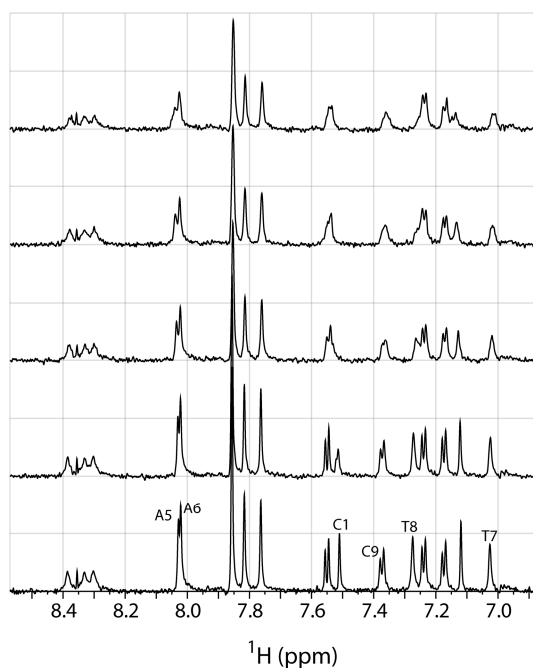


Figure 3.4: Expanded region of the ^1H -NMR titration of $\text{d}(\text{CGCGAATTCGCG})_2$ with derivative **1** (700 MHz, $T = 25\text{ }^\circ\text{C}$). 1 mM (4 mM single strand concentration), in 0.2 ml ($\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1) buffer solution having 10 mM KH_2PO_4 , 70 mM KCl, 0.2 mM EDTA, pH 7.0. Equivalents of the drug are reported on the left of each spectrum. Residues that are slightly affected by the binding are indicated.

Isothermal Titration Calorimetry. ITC is an useful methodology for a complete understanding of drug-DNA interactions, and it has been applied many times to determine the thermodynamic properties of drug-quadruplex interactions. Indeed, ITC is the only technique that directly measures the binding enthalpy change for the formation of a complex, allowing the free energy change to be dissected into the enthalpic and entropic contribution to the association process. This reveals the nature of the forces that drive the binding and can provide insight into the nature of the intermolecular contacts formed and even into changes in solvation⁹. The understanding of those factors can be helpful in both screening among various drugs and optimizing the drug-target interactions, to direct the design of new drugs. ITC experiments reveal that both compound **1** and netropsin bind to the investigated quadruplex. Examples of the raw ITC and integrated heat

data for the titration of $[d(TGGGGT)]_4$ quadruplex with netropsin and compound **1** are shown in Figure 3.5. The ITC data for drugs binding to $[d(TGGGGT)]_4$ indicate, in both cases, the formation of a 2:1 (drug:quadruplex) complex with good affinity. The raw data for the titration of drugs with $[d(TGGGGT)]_4$ (insets in Figure 3.5) indicate an exothermic interaction, based on the positive values observed for the peaks. With each injection of ligand, less and less heat release was observed until constant values were obtained, reflecting, in both cases, a saturable process. The thermodynamic results obtained from fitting the ITC data for netropsin and compound **1** binding to $[d(TGGGGT)]_4$ are given in Table 3.1.

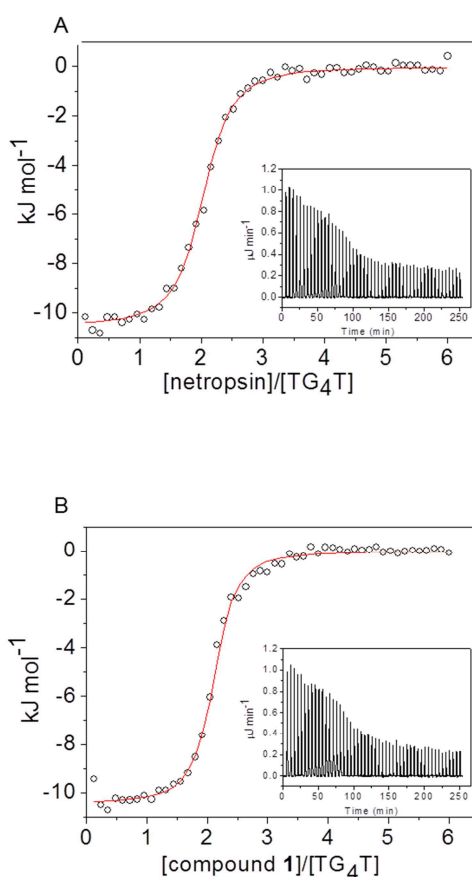


Figure 3.5: Raw ITC data (insets) and binding isotherms for titration of $[d(TGGGGT)]_4$ with netropsin (A) and with compound **1** (B).

Ligand	n	K_b ($\times 10^6 \text{ M}^{-1}$)	$\Delta_b H^\circ$ (kJ mol^{-1})	$T\Delta_b S^\circ$ (kJ mol^{-1})	$\Delta_b G^\circ$ (kJ mol^{-1})
compound 1	2.0 ± 0.1	1.9 ± 0.2	-11.0 ± 2.0	25 ± 2	-36 ± 2
netropsin	2.0 ± 0.1	1.2 ± 0.1	-10.6 ± 1.0	23 ± 2	-34 ± 2

Table 3.1: Thermodynamic parameters for the interaction of compound **1** and netropsin with [d(TGGGGT)]₄ determined by ITC at 25 °C and pH 7.0

The values of the binding constants and the Gibbs energy changes (-34 kJ mol^{-1} and -36 kJ mol^{-1} for netropsin and compound **1**, resp.) indicate that, from a thermodynamic point of view, the interactions with the quadruplex molecule are favoured at 25 °C. The values of ΔH° and $T\Delta S^\circ$ show that in both cases the interactions are associated with a favourable binding enthalpy (-10.6 kJmol^{-1} and -11.0 kJmol^{-1} for netropsin and compound **1**, resp.), however, the binding processes are always entropically driven (23 kJmol^{-1} and 25 kJmol^{-1} for netropsin and compound **1**, resp.). In the case of the interaction of distamycin A with [d(TGGGGT)]₄, similar results were obtained (entropically driven process with a small favourable enthalpic contribution), except for stoichiometry and magnitude of binding constant.

As reported in the literature many times, a groove binding is generally characterized by a large favourable increase of entropy and a small favourable or unfavourable enthalpy change. This could be mainly due to the fact that the association reaction is driven in large part by a hydrophobic effect and the favorable entropy is derived from the release of bound water molecules from the DNA and drug upon complex formation as well as release of counterions upon binding of cationic ligands. Interestingly, the entropically driven interactions of the compound **1** and of netropsin suggest a groove-binding mechanism also in these cases, as

observed for distamycin A. Additionally, in contrast to many other groove binders, netropsin binding to duplex DNA is accompanied on average by a favourable enthalpy change. This is the result of both electrostatic interactions and hydrogen bonds formation between netropsin molecule and DNA. A favourable enthalpy change is observed by ITC for netropsin binding to [d(TGGGGT)]₄ as well as for compound **1**, suggesting once again a similar groove binding mode.

As stated before, the stoichiometry obtained for the association of the investigated ligands to [d(TGGGGT)]₄ is 2:1 (drug : quadruplex), while for distamycin A a stoichiometry of 4:1 was found. This finding is consistent with the observation that the distamycin A molecules are able to bind as dimers, while netropsin (and probably also compound **1**) is not able to dimerise because of charge repulsions. Notably, in the distamycin A-[d(TGGGGT)]₄ complex, each distamycin A dimer expands its bounded groove (similarly to that observed with duplex DNA), with concomitant reduction of the size of the adjacent ones, preventing a further interaction with other ligand molecules. In the binding of netropsin to duplex DNA the minor groove is widened by 0.5–2.0 Å by the entry of the drug⁶. We can speculate that the binding of the two netropsin molecules to [d(TGGGGT)]₄ could similarly expand the bounded grooves, reducing the size of the adjacent ones and preventing more interactions.

Finally, we also performed experiments on the interaction of compound **1** with the DNA duplex d(CGCGAATTCGCG)₂ containing the central AATT core, which is considered the specific binding site for distamycin A and netropsin, to evaluate if the structural modifications of compound **1** influence the binding. An example of

the raw ITC data for the titration of the d(CGCGAATTCGCG)₂ duplex with compound **1** is shown in Figure 3.6.

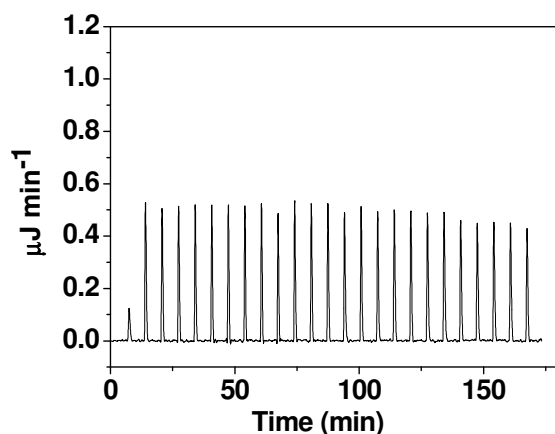


Figure 3.6: Raw ITC data for titration of d(CGCGAATTCGCG)₂ duplex with compound **1**.

Resolvable binding isotherm was never obtained for the interaction of compound **1** with duplex using any combination of reactant concentrations, suggesting low affinity of the molecule for the investigated duplex. On the contrary it is well known that both distamycin A and netropsin have a high binding constant (between 10^6 and 10^8M^{-1}) for that target containing the specific binding site. This finding demonstrates that the structural modifications of compound **1** decrease the affinity of the ligand toward the duplex, enhancing the selectivity.

3.3 Conclusions

In summary The combination of calorimetric and NMR methodologies has enabled us to demonstrate that netropsin and compound **1** are able to bind to [d(TGGGGT)]₄ with good affinity, forming, in both cases, a 2:1 (drug : quadruplex)

complex. Very interestingly, this study shows that the structural modifications of compound **1** do not influence, compared to netropsin, the interaction with the quadruplex, but such modifications decrease the affinity of the ligand toward the duplex, enhancing the selectivity.

3.4 Materials and Methods

Materials. The d(TGGGGT) and d(CGCGAATTCGCG) oligonucleotide sequences used for this study were purchased from the Primm Company (Milan, Italy). Quadruplex and duplex samples were prepared by dissolving the lyophilised compound in a buffer solution containing 20 mM phosphate with 70 mM KCl, 0.1 mM EDTA at pH 7.0. The resulting solutions were annealed by heating at 95 °C for 5 minutes. The solutions were then slowly cooled to room temperature and equilibrated for 1 day at 4 °C. The concentration of oligonucleotides was determined by UV adsorption measurements at 90 °C using molar extinction coefficient values $\epsilon_{(260\text{ nm})}$ of 57800 and 110700 M⁻¹ cm⁻¹ for d(TGGGGT) and d(CGCGAATTCGCG), respectively. The molar extinction coefficients were calculated by the nearest neighbour model¹³. Compound **1** has been synthesized as reported in the literature¹⁴, while netropsin has been purchased from Sigma-Aldrich (St. Louis, MO). Drug solutions have been prepared in the same buffer used for the oligonucleotides, and the concentration has been estimated by UV spectroscopy using the calculated extinction coefficient value $\epsilon_{(297\text{ nm})}$ of 30000 M⁻¹ cm⁻¹ for compound **1** and the reported extinction coefficient value $\epsilon_{(296\text{ nm})}$ of 21500 M⁻¹ cm⁻¹ for netropsin¹⁵.

Nuclear Magnetic Resonance Experiments. [d(TGGGGT)]₄ and d(CGCGAATTCGCG)₂ NMR samples were prepared at a concentration of 2 mM and 1 mM respectively, in 0.2 ml (H₂O/D₂O 9:1) buffer solution having 10 mM KH₂PO₄/K₂HPO₄, 70 mM KCl, 0.1 mM EDTA, and pH 7.0. NMR spectra were recorded with Varian ^{Unity}INOVA 700 MHz spectrometer. ¹H chemical shifts were referenced relative to external sodium 2,2-dimethyl-2-silapentane-5- sulfonate (DSS). 1D proton spectra of samples were recorded using pulsed-field gradient DPGSE for H₂O suppression^{16,17}.

Isothermal Titration Calorimetry Experiments. ITC experiments were carried out at 298 K using a high sensitivity CSC-5300 Nano-ITC microcalorimeter from Calorimetry Science Corporation (Lindon, Utah) with a cell volume of 1 ml. Before each ITC experiment, the pH of each solution was checked, the reference cell was filled with deionised water, and the DNA solutions were degassed for 5 minutes to eliminate air bubbles. Care was taken to start the first addition after baseline stability had been achieved. In each titration, volumes of 5–10 µL of a solution containing compound **1** or netropsin at a concentration of 600–700 µM were injected into a solution of quadruplex or duplex DNA (30 µM) in the same buffer, using a computer-controlled 250 µL microsyringe. In order to allow the system to reach the equilibrium, we applied a spacing of 300 or 400 s between each ligand injection. Heat produced by ligands dilution was evaluated by performing a control experiment, titrating each ligand into the buffer alone. The interaction heat for each injection was calculated after correction for the heat of ligand dilution. The corrected heat values were plotted as a function of the molar ratio, to give the

corresponding binding isotherms. The resulting isotherms were fitted to a single set of identical sites model employing a nonlinear least-squares minimisation algorithm to a theoretical titration curve, using the program Bindwork from Calorimetry Science Inc. ΔH° (reaction enthalpy change in kJmol^{-1}), K_b (binding constant in M^{-1}), and n (number of binding sites) were the fitting parameters. The Gibbs energy and the entropic contribution were calculated using the relationships $\Delta_b G^\circ = -RT \ln K_b$, ($R = 8.314 \text{ Jmol}^{-1} \text{ K}^{-1}$, $T = 298 \text{ K}$) and $-T\Delta_b S^\circ = \Delta_b G^\circ - \Delta_b H^\circ$.

References

- [1] Haq I. *Arch. Biochem. Biophys.* **2002**,. 403, 1-15.
- [2] Cantor C. R., Warshaw M. M., Shapiro H. *Biopolymers*, **1970**, 9, 1059-1077.
- [3] Di Pietro G., Giannini G., Iafrate E. M. *et al. J. Chem. Soc.*, **1996**, Perkin Trans. 1, 1333-1335.
- [4] Hwang T. L. and Shaka A. J. *J. Magn. Res.*, **1995**, A112, 275-279.
- [5] Wartell R. M., Larson J. E., Wells R. D. *J. Biol. Chem.*, **1974**, 249, 6719-6731.
- [6] Randazzo A., Galeone A., Mayol L. *Chem. Commun.*, **2001**, 11, pp. 1030-1031.
- [7] Cocco M. J., Hanakahi L. A., Huber M. D. and Maizels N. *Nucleic Acids Res.*, **2003**, 31, 2944-2951.
- [8] Pagano B., Mattia C. A. and Giancola C. *Int. J. Mol. Sci.*, **2009**, 10, 2935-2957.
- [9] Chaires J. B. *Arch. Biochem. Biophys.* , **2006**, 453, 24-29.
- [10] Pagano B., Virno A., Mattia C. A., Mayol L., Randazzo A., Giancola C. *Biochimie*, **2008**, 90, 1224-1232.
- [11] Chaires J. B. *Arch. Biochem. Biophys.* **2006**, 453, 24-29.
- [12] Zimmer C. *Prog. Nucleic Acid Res. Mol. Biol.*, **1975**, 15, 285-318.
- [13] Dalvit C. *J. Biomol. NMR*, **1998**, 11, 437-444.
- [14] Hare D. R., Wemmer D. E., Chou S. and Drobny G. *J. Mol. Biol.*, **1983**, 171, 319-336.
- [15] Wartell R. M., Larson J. E. and Wells R. D. *J. Biol. Chem.*, **1974**, 249, 6719-6731.
- [16] Rentzeperis D., Marky L. A., Dwyer T. J., Geierstanger B. H., Pelton J. G. and Wemmer D. E. *Biochemistry*, **1995**, 34, 2937-2945.
- [17] Pelton J. G., Wemmer D. E. *J. Am. Chem. Soc.*, **1990**, 112, 1393-1399.

4) Structural and conformational requisites in DNA quadruplex groove binding: another piece to the puzzle

4.1 Introduction

Here, we report a distamycin A (**3**) analogue (Figure 4.1), where the major change into the structure is the replacement of amidinium group by an uncharged N-methyl amide moiety, to probe the real importance of the unique Coulombic interaction in the distamycin A/[d(TGGGGT)]₄ complex formation. Particularly, the synthesis, the NMR titration profile, and the three-dimensional structure of **3**/[d(TGGGGT)]₄ complex are reported. To thoroughly characterize the binding of the distamycin A derivative with the G-quadruplex DNA motif, a calorimetric study was also performed.

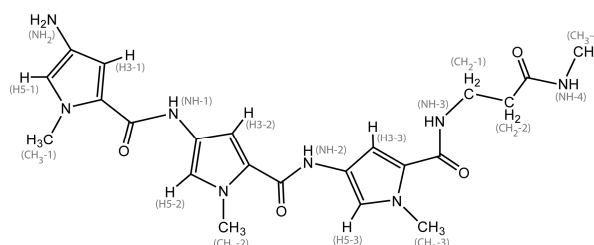


Figure 4.1: Chemical structure of derivative **3**.

4.2 Results and Discussion

Synthesis of Derivative 3. The novel derivative **3** was obtained following the procedure reported in the Scheme 1. Starting from derivative **1**, prepared by a methodology previously reported¹, it was condensed with 3-amino-N-

methylpropanamide hydrochloride² using an excess (2 equiv.) of 1-ethyl-3-[3-(dimethylamino) propyl]carbodiimide hydrochloride (EDCI) as coupling agent, in DMF and in presence of Hunig's base at room temperature for 18 h, to obtain the N-methyl carbamoyl derivative **2** in good yield after purification by silica gel flashchromatography.

The amido modified distamycin A derivative **2** was then transformed in the corresponding desformyl derivative **3** by treatment of HCl in a mixture of ethanol and water. The reason for this last synthetic step was to obtain the hydrochloride derivative **3**, that turned out to be more soluble in pure H₂O than derivative **2** at the concentration used for the preparation of the NMR samples.

Nuclear Magnetic Resonance Experiments. To evaluate the binding properties of derivative **3**, and to perform a direct comparison with the binding behavior of distamycin A, [d(TGGGGT)]₄ has been titrated with **3** at the same experimental conditions (buffer, temperature, DNA concentration) used for distamycin A³. As far as distamycin A is concerned, below 2:1 ligand/quadruplex stoichiometry, the addition of distamycin A to [d(TGGGGT)]₄ caused gradual changing in chemical shift of the signal of the quadruplex, whereas further addition of drug caused the appearance of a new set of proton signals, whose intensities rose by increasing the amount of drug with the concomitant falling off of the original signals which completely disappeared at a ratio of 4:1 drug-DNA.

On the other hand, the NMR titration profile with **3** turned out to be clearly different from that observed in the case of distamycin A (Figure 4.2).

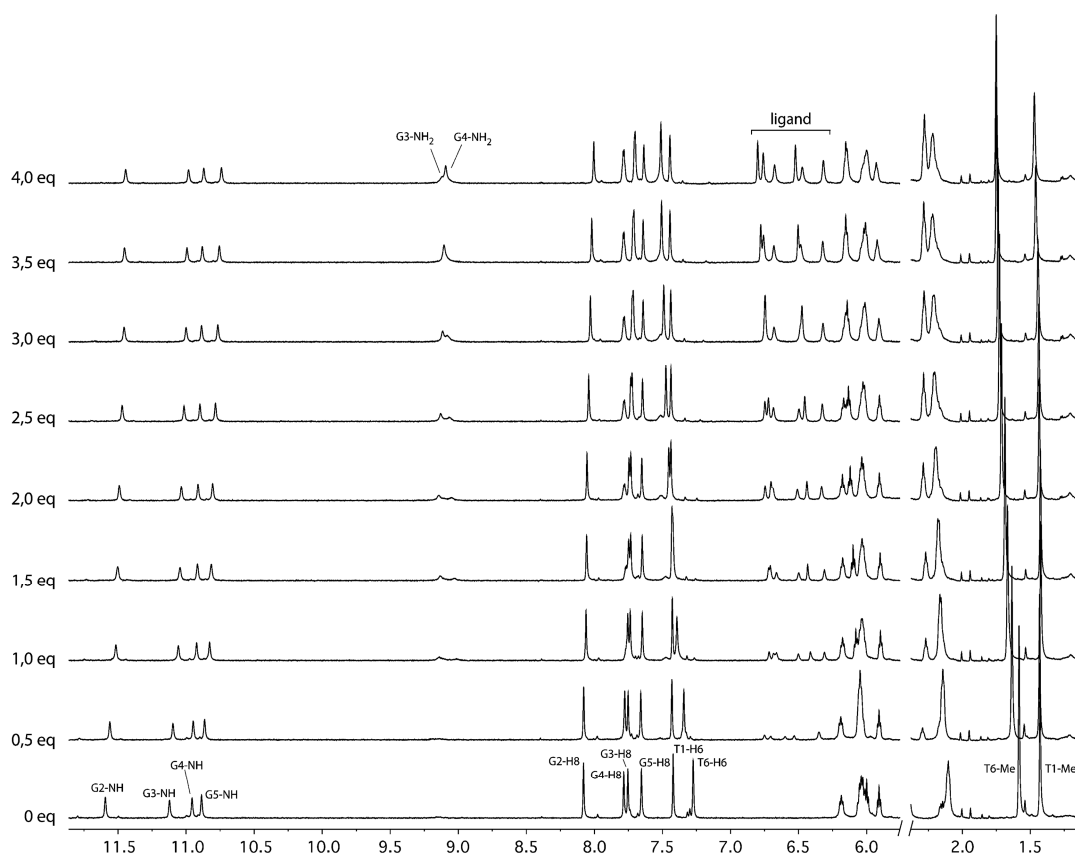


Figure 4.2: NMR titration of $[d(TGGGGT)]_4$ with derivative **3** (700 MHz, $T = 25\text{ }^{\circ}\text{C}$). 2 mM (8 mM single strand concentration), in 0.6 ml ($\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1) buffer solution having 10 mM KH_2PO_4 , 70 mM KCl, 0.2 mM EDTA, pH 7.0. Equivalents of the drug are reported on the left of each spectrum.

In fact, the addition of **3** to $[d(TGGGGT)]_4$ caused only gradual changes in the chemical shift of DNA proton resonances. At ligand/DNA ratio of 4:1, the titration was virtually completed. The four strands resulted to be magnetically equivalent throughout the titration, and no splitting of resonances was observed at any stage. Moreover, it is noteworthy that, throughout the whole NMR titration, a single set of signals was present for derivative **3** protons, which only grew in intensity and did not show any significant change in chemical shift values by increasing ligand concentration. These observations suggest that derivative **3** bind the quadruplex in a fast process on the NMR time scale.

To preliminarily evaluate the binding site of derivative **3**, a comparison of resonances of some protons of the uncomplexed DNA and the complexed one has been done. In particular, we report the $\Delta\delta$ values (chemical shifts of the complex minus free DNA) of aromatic, methyl and imino protons in Figure 4.3A. Interestingly, all the analyzed resonances shifted. Nevertheless, the signal of the protons of T1 residue shifted the least, whereas the ones of residue T6 the more. In any case, a general shift of the aromatic and imino signals was observed also for the G2, G3, G4, and G5. This means that, basically, derivative **3** is able to recognize most of the molecule, even if it recognizes preferentially the 3' edge of the quadruplex.

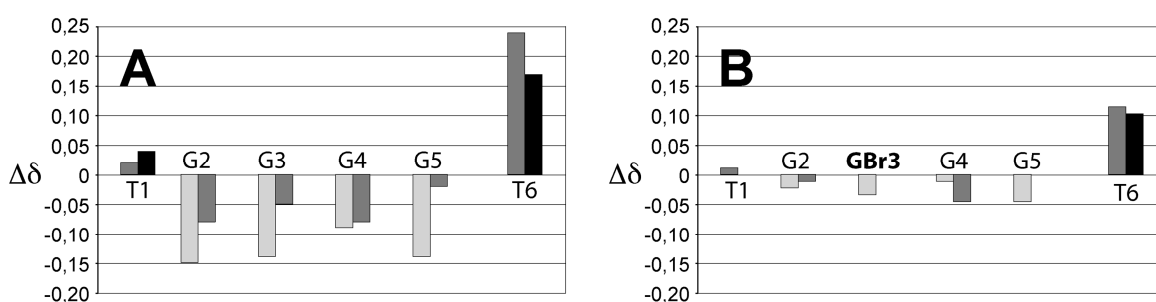


Figure 4.3: Difference in chemical shifts ($\Delta\delta$) of DNA upon binding of **3** (complex **3**:DNA minus DNA alone) to [d(TGGGGT)]₄ (A) and [d(TGG^{Br}GGT)]₄ (B). $\Delta\delta$ values are reported for imino (light gray), aromatic (dark gray) and methyl (black) protons. No aromatic hydrogens are present for brominated residues.

An almost complete assignment of the nonexchangeable/exchangeable protons of the complex has been accomplished by means of a combination of the analysis of 2D NOESY and TOCSY spectra (700 MHz, T = 25 °C). In particular, as far as DNA is concerned, ¹H resonances within each deoxyribose were identified by 2D TOCSY experiment, while the analysis of NOEs among base protons and H1', H2', and H2'' protons allowed us to assign all base protons. The direct comparison of the intensities of the NOESY crosspeaks (700 MHz, T = 25 °C, mixing time 100 ms)

between the H8 proton bases and sugar H1' resonances, and among H8 proton bases and sugar H2'/H2'' resonances, indicates that all Gs residues of the complexed DNA adopt an *anti* glycosidic conformation. Then, all bases have classical H8/H2'-H2'' sequential connectivities to 5' neighboring, indicating that the four strands are involved in the formation of a helical structure. Moreover, the entire pattern of NOEs observed indicates that the backbone conformation resembles closely that of the uncomplexed [d(TGGGGT)]₄ possessing a right-handed B-form helix structure.

Each exchangeable proton signal was then assigned to the pertinent hydrogen by an in-depth analysis of the NOESY spectra. In particular, the NOE contact between H1 proton of DNA at δ H 11.34 and the methyl protons at δ H 1.34 of the T1 residue led us to identify the imino protons belonging to the tetrad that is in proximity of the 5' edge of the quadruplex. Similarly, the identification of the imino protons belonging to the tetrad in proximity of the 3' edge was confirmed by the NOE between H1 proton at δ H 10.64 and the methyl group at δ H 1.66 ppm of the thymine T6. The other four H1 resonances of the remaining two central tetrads were identified by analyzing the NOE connections with the adjacent tetrads.

The resonances of derivative **3** were assigned, first, identifying the resonances belonging to the hydrogen on the convex side of the molecule. Particularly, the H5 pyrrole protons (δ H 6.70, 6.66, and 6.57 ppm for H5-1, H5-2 and H5-3, respectively) have been assigned taking into account that they both exhibit scalar and dipolar coupling with methyls that are linked to the same pyrrole ring (δ H 3.52, 3.40, and 3.37 ppm for CH₃-1, CH₃-2, CH₃-3, respectively). As for the protons on the concave side of derivative **3**, the H3 pyrrole protons (δ H 6.42, 6.37, and 6.22 for H3-1, H3-2 and H3-3, respectively) have been assigned by virtue of very small

scalar coupling with H5 protons of the same pyrrole ring. Unfortunately, we were not able to unequivocally assign NH-1, NH-2 and NH-3 protons due to the presence of ambiguous NOEs with the adjacent H3s.

The NOESY spectrum of the complex contains a number of intermolecular ligand-ligand and ligand-DNA NOEs, in addition to intramolecular ones. Particularly, as for ligand ligand contacts, 12 head to tail NOEs (for each molecule of **3**, 24 for each dimer) were clearly discernible in the NOESY spectra (Table 4.1), that unambiguously indicates that the ligand molecules bind to the quadruplex, two by two, with each term of the dimeric pairs with an antiparallel orientation and in close contact to its partner (as observed for the binding of distamycin A to both duplex and quadruplex DNA)^{3,4}. In addition, 14 NOEs were observed between each molecule of derivative **3** and [d(TGGGGT)]₄. In particular, derivative **3** displayed NOE contacts predominantly with residues G4, G5, and T6 of DNA. Analogously to that observed for distamycin A, also in this case, there are no contacts between derivative **3** and the imino protons of the quadruplex (since these point inside the quadruplex core) (Table 4.2). All this suggests that derivative **3** actually does not stack on the edges of the quadruplex, even if it prefers to interact with the 3' edge of the quadruplex.

Ligand Proton	Ligand Proton
H3-1	CH ₃ -4, CH ₃ -3, H5-3
H5-1	CH ₂ -1, CH ₃ -4, CH ₃ -3
H3-2	CH ₃ -3, H3-3, H5-3
H5-2	CH ₃ -4
H3-3	CH ₃ -1
H5-3	CH ₃ -1

Table 4.1: Intermolecular head to tail ligand ligand NOE contacts

Ligand Proton	DNA Proton
H3-1	T6:H2', T6:H1', G5:H2', G3:NH ₂
H5-1	G5:H1', G5:H2', G4:H2'
CH ₃ -1	T6:H1'
H3-2	T6:H1'
H5-2	G5:H2'
H3-3	T6:H6, G5:H1', G4:H2'
H5-3	G4:H2', G5:H1'

Table 4.2: Intermolecular ligand-DNA NOE contacts

Synthesis and NMR Titration of the Quadruplexes Containing Bromine

Atoms. To further confirm that **3** is able to bind the grooves of the quadruplex, as its analogue distamycin A, we have designed and synthesized a modified oligonucleotide, namely, d(TGG^{Br}GGT), where dG^{Br} is 8-bromo-2'-deoxyguanosine, potentially capable to form quadruplex structure and possessing a bulky group (bromine) at the very center of the grooves.

Our intention was to titrate d(TGG^{Br}GGT) with **3** and to analyze the titration profile. As a result, if **3** interacts with the portion of the grooves of the quadruplex where the bromine is present, this should prevent (or at least should limit) the insertion of molecules of **3** into the grooves. However, we first tested the capability of d(TGG^{Br}GGT) to form a quadruplex structure. NMR sample of d(TGG^{Br}GGT) was prepared at a concentration of 2 mM, in 0.6 mL (H₂O/D₂O 9:1) buffer solution having 10 mM KH₂PO₄, 70 mM KCl, 0.2 mM EDTA, pH 7.0. The sample was annealed for 5-10 min at 80 °C and slowly cooled down to room temperature, then ¹H-NMR spectra were recorded using DPGSE pulse sequence for H₂O suppression^{5,6}.

The ¹H-NMR spectrum (700 MHz, T = 25 °C) of d(TGG^{Br}GGT) shows the presence of four well-defined singlets in the region 11-12 ppm (data not shown), attributable to imino protons involved in Hoogsteen hydrogen bonds of G-quartets, and the presence of five signals (Figure 4.4), belonging to three guanine H8 and two

thymine H6 protons in the aromatic region. These indicate that a single well defined quadruplex species is present in solution, consisting of four G-tetrads and possessing a 4-fold symmetry with all strands parallel to each other. In a quadruplex like this, each Br group of the quadruplex faces right into the grooves, pointing outward the quadruplex. Thus, the quadruplex $[d(TGG^{Br}GGT)]_4$ has been titrated with **3** and the results are reported in Figure 4.4.

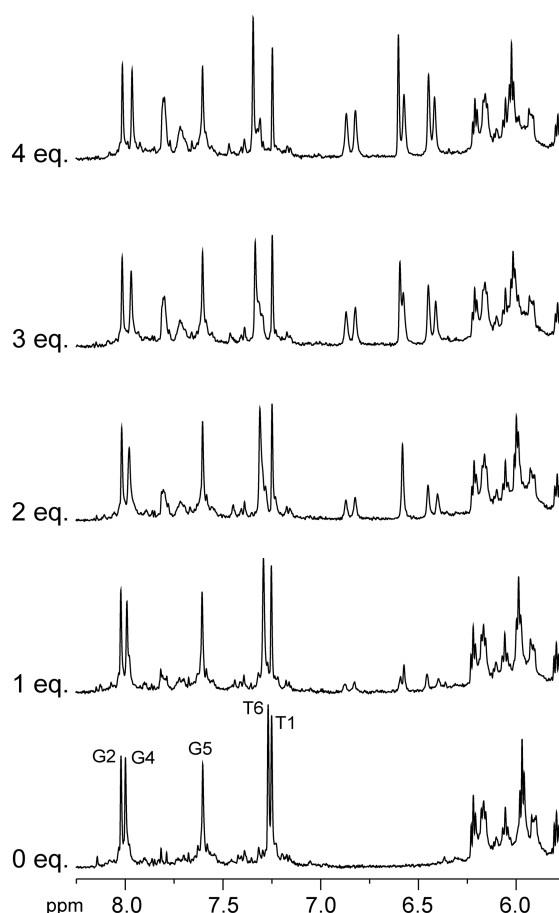


Figure 4.4: Expanded regions of NMR titration of $[d(TGG^{Br}GGT)]_4$ with derivative **3** (700 MHz, T = 25 °C). 2 mM (8 mM single strand concentration), in 0.6 ml (H_2O/D_2O 9:1) buffer solution having 10 mM KH_2PO_4 , 70 mM KCl, 0.2 mM EDTA, pH 7.0. Equivalents of the drug are reported on the left.

The progressive increase of concentration of **3** up to 4 mol equiv caused only a slight drift of DNA signals and drug resonances to gradually grow in intensity

(between 6 and 7 ppm). Further additions of drug did not lead to significant changes. As suggested also by the $\Delta\delta$ values of DNA signals (Figure 4.3B), the extent of chemical shift perturbation is lighter in comparison to the unmodified quadruplex. This strongly indicates that, actually, the presence of the bromines at the very center of the grooves does affect the recognition process, and that **3** binds the grooves of [d(TGGGGT)]₄. This also suggests that, in spite of **3** which does not show any NOE contacts with the 5' edge of the groove, **3** is actually able to span the entire groove.

Structure Calculations. To obtain the three-dimensional structure of the 4:1 complex at atomic level, an estimation of proton-proton distances has been retrieved from cross-peak intensities in 2D NOESY experiments (700 MHz, T = 25 °C). A total of 278 experimental distance restraints were used for the calculations, and as suggested by the presence of eight guanine imino protons in the 1D ¹H-NMR spectrum, 32 supplementary distance restraints (HN1-O6, HN2-N7) for 32 hydrogen bonds corresponding to the four G-quartets were also incorporated during the computations (Table 4.3).

Restraints for a number of backbone torsion angles were also used in the calculations. In particular, the combined analysis of the PE-COSY and zTOCSY experiments revealed that all measurable $J_{H1',H2'}$ and $J_{H1',H2''}$ ranged from 7 to 9 Hz and from 5 to 6.5 Hz, respectively, and that a number of crosspeaks for H3'/H4' correlations were missing. Furthermore, the measurable $J_{H3',H4'}$ values were very low. All this indicates a predominant S-type nature of sugar ring conformations.

Total no. of experimental NOE restraints	278
Quadruplex intraresidual restraints	96
Quadruplex interresidual restraints	68
Intermolecular ligand-quadruplex restraints	62
Intermolecular ligand-ligand restraints	52
Dihedral angles (α , β , γ , δ , ϵ and χ) restraints	96
Quadruplex H-bonds restraints	32
Planarity restraints for G bases	56
Average rmsd (all atoms excluding terminal T bases)	0.84 Å
<i>Energy (kcal/mol)</i>	
Mean amber energy	-1610.42 \pm 62.29
NOE distance restraints violation energy	84.34 \pm 4.82
Torsion angle restraints violation energy	184.10 \pm 18.29
<i>Restraint violations</i>	
Distance (> 0.3 Å)	7
Experimental dihedral angles (> 10 Å)	2

Table 4.3: Structural Data for the NMR Restraints and statistics for the 10 best structures

Therefore, the relative δ angles constraints have been consistently constrained into the range $95^\circ/175^\circ$ ⁷. Very interesting information could be also retrieved from the analysis of the sum of H4' couplings constant ($\Sigma J_{H4'} = J_{H4',H3'} + J_{H4',H5'} + J_{H4',H5''}$). All measured $J_{H4'}$ turned out to be 10 Hz. This, along with the lack of intense NOEs between H8 and any H5'/H5''protons, suggested that γ angles could be constrained in the range $20^\circ/100^\circ$ ⁷. As far as θ and ϵ angles are concerned, their estimation have been performed measuring the scalar coupling in the 2D proton-detected heteronuclear ^1H - ^{31}P COSY and by using the semiempirical Karplus equation⁷. Thus, since for all $J_{P,H5'}$ and $J_{P,H5''}$ were 8 Hz, θ backbone torsion angles were restricted to $-230^\circ/-110^\circ$ ⁷. On the contrary, $J_{P,H3'}$ for residues 2, 3, 4, and 5 were larger than 8 Hz, and the corresponding ϵ angles were restricted to $-230^\circ/-$

110°⁷. Furthermore, since α (O3'-P-O5'-C5') and ζ (C3'-O3'-P-O5') dihedral angles do not involve any protons, it has been impossible to directly retrieve structural information for these angles. Nevertheless, since in a structured DNA α and ζ angles cannot be considered independent from β , γ , δ and ϵ , we thought opportune to restraint at least one of these angles; thus, α angles have been constrained in the wide range -150°/-30°. Finally, due to the presence of very weak NOEs between H8 and H1' hydrogens, glycosidic torsion angles χ were fixed in the *anti*-domain (-155°/-75°).

Therefore, 3D structures which satisfy NOEs and dihedral angles were constructed by simulated annealing (SA) calculations. A total of 100 structures was generated, and among them, the 10 best models were selected based on the value of the overall potential energy and NMR restraint violations for further simulations. To probe the thermodynamic stability of these structures and to further refine the models, these were all subjected to a 5 ns MD simulation in explicit solvent. All the 10 MD runs demonstrated a good stability of the whole structure. Moreover, the aforementioned simulations converged toward a well-defined ligand/DNA conformation as depicted in Figure 4.5 reporting the representative conformations (namely, the ones with the lowest rmsd value with respect to the average structure calculated over the whole production run).

As expected, differently from what demonstrated for distamycin A, compound **3** dimers are shifted toward the 3' end of the quadruplex. The two staggered antiparallel **3** molecules overlap for their total length with each N-methylpyrrole ring of one molecule contacting the terminal amide bond of the other, thus, making synchronized twisting to easily fit the curvature of the DNA

quadruplex grooves. Moreover, one molecule of the dimer of **3** is more solvent exposed, while the other is sandwiched between the first one and the quadruplex structure (herein referred as **3**_{ex} and **3**_{in}, respectively). The preference for such an orientation with respect to the G-quadruplex could be ascribable to the absence of a positively charged anchoring point that in distamycin A stabilizes each monomer so as to project the methyl groups toward the solvent and the opposite amide protons toward the crevice of the groove.

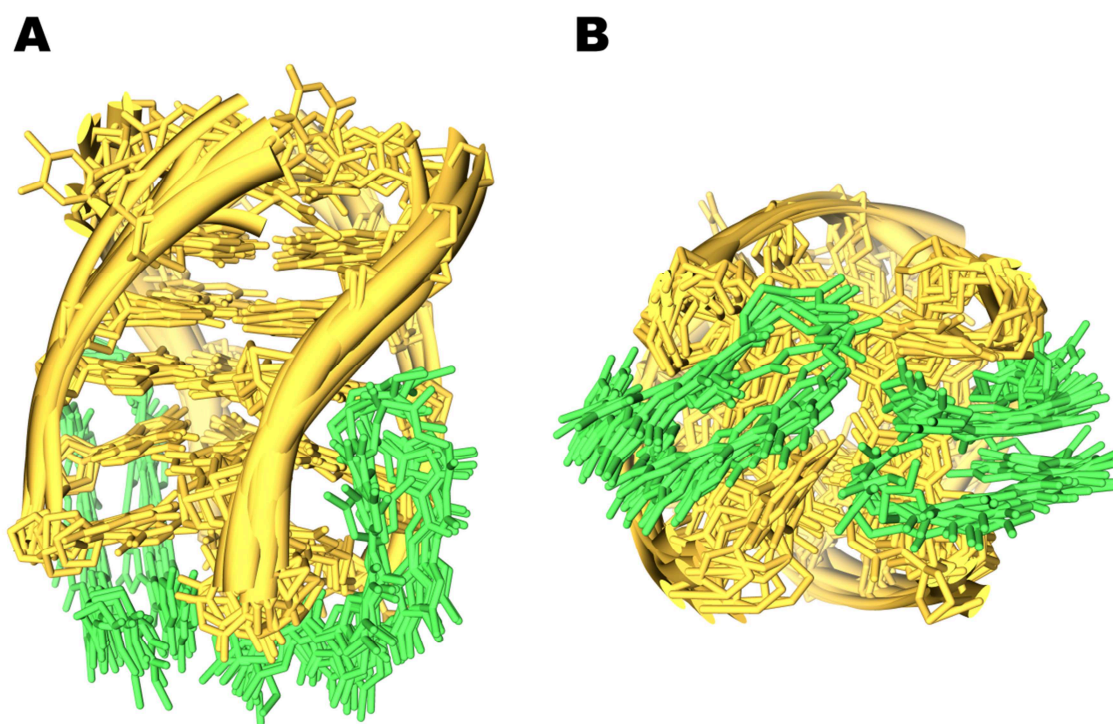


Figure 4.5: Side view (**A**) and bottom view (**B**) of the superimposition of the 10 best structures of the 4:1 **3**/[d(TGGGGT)]₄ complex. Derivative **3** is reported in green, DNA is colored in yellow.

As aforementioned, the two molecules of compound **3** interact with the quadruplex in an antiparallel fashion, as its analogue distamycin A. This is a very interesting result; in fact, in theory, two possible orientations could be predicted for a dimer of **3**: parallel and antiparallel. This behavior could be imputed to the charge-

transfer π - π interactions and to the energetically favorable dipole-dipole interactions established between the monomers as demonstrated by the dipole moments of the two staggered structures which are pointing in opposite directions.

The crescent shape of the ligand allows the formation of favorable interactions with the quadruplex by following the curvature of the grooves in the G4-G5 region and taking contact with T6 at the 3' end. A comparison with the experimental distamycin A/[d(TGGGGT)]₄ complex reveals that in both cases a crescent shape is adopted by the ligand. Nevertheless, such a conformation alone does not seem to guarantee a pure groove binding mode in the [d(TGGGGT)]₄. This indicates that the concurrence of the crucial anchoring positively charged point (amidinium group) and the crescent shape conformation in distamycin A are both responsible for a pure groove binding mode.

Calculation of the Delphi electrostatic potential surface reveals that the terminal methylcarboxamide groups of the **3** dimer are hosted in rather polar regions of the quadruplex while the terminal pyrrole ring is exposed to the solvent in the G4 region in **3**_{ex} or embedded in a less hydrophilic environment in the T6 region in **3**_{in}.

A close-up view of the best binding pose resulting from NMR constrained SA experiments reveals that the aforementioned terminal groups of **3** are involved in a network of H-bonds with the quadruplex (Figure 4.6).

Most precisely, the flexible carboxamide terminal group of the **3**_{in} molecule is kept in place by an intramolecular H-bond between its NH-4 proton and the carbonyl oxygen attached to the third pyrrole ring and by an intermolecular one established between the terminal carbonyl oxygen of **3**_{in} and the terminal NH₂

group of **3**_{ex}. Moreover, the G5 base is contacted by the dimer through two H-bonds between its O4' and its N3 atoms with the **3**_{ex} NH₂ and **3**_{in} NH-4 groups, respectively. While polar contacts mainly govern the recognition between the **3** dimer and the center of the quadruplex groove, rather hydrophobic interactions are established with the 3' end of the [d(TGGGGT)]₄ structure. Notably, the terminal pyrrole ring of the **3**_{in} molecule is embedded in a sort of aromatic cage formed by the two T6 bases of the groove and the third pyrrole of the **3**_{ex} molecule.

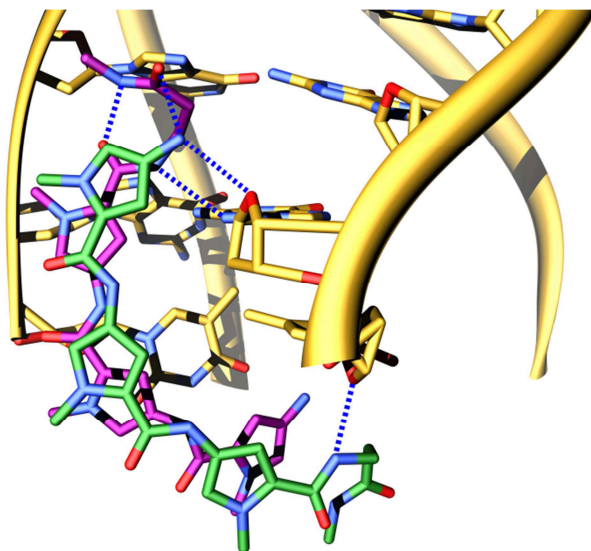


Figure 4.6: Side view representation of the 4:1 complex **3**/[d(TGGGGT)]₄. **3**_{in}, **3**_{ex} and [d(TGGGGT)]₄ are depicted in magenta, green and yellow sticks, respectively.

Isothermal Titration Calorimetry Measurements. Thanks to the great advances in the sensitivity and reliability of the calorimeter, ITC has become an important tool for the direct measurement of thermodynamic parameters in several biological interactions. Moreover, it has been proven to be useful to study the energetic aspects of interactions between G-quadruplexes and other molecules, including small ligands⁸.

Therefore, we have utilized ITC to characterize the thermodynamics of binding of derivative **3** to [d(TGGGGT)]₄ quadruplex. Figure 4.7 shows the binding isotherm resulting from the integration of raw data after correction for the heat of ligand dilution.

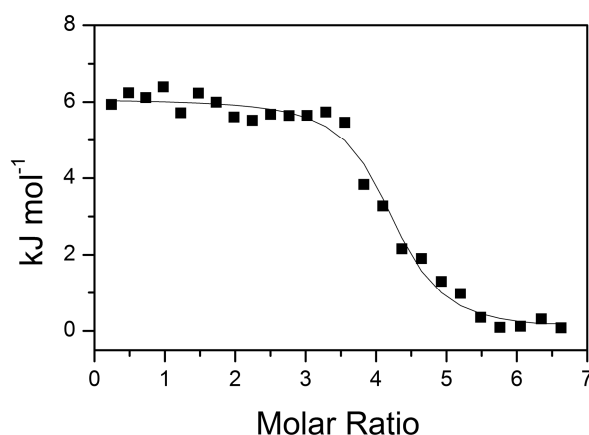


Figure 4.7: Normalized heat of interaction between compound **3** and [d(TGGGGT)]₄. The squares represent the experimental data obtained by integrating the raw data and subtracting the heat of ligand dilution into the buffer. The line represent the best fit obtained by a nonlinear least-squares procedure based on a single set of identical sites model.

The calorimetric data indicate an endothermic interaction. After a number of injections of ligand, increasingly less heat uptake was observed until constant values were obtained (corresponding to the heat of dilution), reflecting a saturable process. The values of K_b and ΔH° derived by ITC enable us to complete the thermodynamic binding profiles by calculating corresponding values of $T\Delta S^\circ$ and ΔG° . The resulting thermodynamic parameters are showed in Table 4.4.

Drug	n	K_b (M ⁻¹)	ΔH° (kJ mol ⁻¹)	$T\Delta S^\circ$ (kJ mol ⁻¹)	ΔG°_{298K} (kJ mol ⁻¹)
3	4	9×10^5	6	40	-34

Table 4.4: Thermodynamic parameters for the interaction of compound **3** with [d(TGGGGT)]₄ determined by ITC at 25 °C

The binding affinity measured from ITC is $9 \times 10^5 \text{ M}^{-1}$ at 25 °C. The binding enthalpy was found to be positive ($\Delta H^\circ = 6 \text{ kJ mol}^{-1}$) and the stoichiometry of interaction was found to be 4:1 ligand-quadruplex. Inspection of the data reveals that the entropic contribution ($T\Delta S^\circ = 40 \text{ kJ mol}^{-1}$) to the binding of derivative **3** to the quadruplex provides the driving force for the ligand-quadruplex binding reaction, while the positive heat of formation of the complex indicates that this interaction is an enthalpically unfavorable process.

The results from the calorimetric titration data alone do not allow us to demonstrate with certainty the binding mode of this ligand. Usually, small molecules bind to duplex DNA by two predominant binding modes, intercalation and groove binding. Calorimetric studies have determined the enthalpic and entropic contributions to the DNA binding of representative DNA binding compounds. Generally, the binding of an intercalator to DNA is driven by a large favorable enthalpy change and by an unfavorable entropy decrease, while the binding of a groove binder to DNA is characterized by a large favorable increase of entropy and a small favorable or unfavorable enthalpy change⁹. In this case, the thermodynamic data determined by ITC for the binding of derivative **3** to [d(TGGGGT)]₄ (large favorable increase of entropy and small unfavorable enthalpy change) are in agreement with the groove binding profile and are confirmed by the structure of the complex.

The thermodynamic parameters for the interaction of derivative **3** are slightly different from the ones determined by ITC for the binding of distamycin A to the same quadruplex molecule. In both cases, the driving force for the binding

reaction is the entropy change. However, the binding of derivative **3** shows a small unfavorable enthalpy change, while in the case of distamycin A, a small favorable enthalpy change was observed. This interesting difference is probably due to structural features of the two ligands. Indeed, the positively charged amidinium moiety of distamycin A interacts with the phosphate groups of the quadruplex, providing a favourable (although small) enthalpy contribution, while in the derivative **3**, the amidinium group is replaced by the uncharged N-methyl amide moiety that cannot give this contribution.

4.3 Conclusions

In summary, the structural data herein reported unveil the role of the Coulombic interaction engaged by distamycin A and the quadruplex. Such an interaction not only influences the strength of ligand/quadruplex formation, but surprisingly governs the orientation of the ligand with respect to the DNA. In fact, the removal of the positively charged terminal group results in an unprecedented ligand binding position in which both the groove and the 3' end of the DNA are occupied. Furthermore, the lack of charge in the ligand does not affect the relative orientation of two molecules of **3** forming a dimer. All this opens up the tempting opportunity for a fine-tuning of the drug-quadruplex interaction mode. Moreover, with this further contribution, we provided the scientific community with an enhanced knowledge of the structural and conformational demands of the quadruplex groove that will serve as a platform for a rational design of novel groove binding agents.

4.4 Materials and Methods

Oligonucleotide synthesis. The oligonucleotide d(TGGGGT) was synthesized on a Millipore Cyclone Plus DNA synthesizer using solid phase β -cyanoethyl phosphoramidite chemistry at 15 μ mol scale. Commercially available 5'-DMT-aminoprotected-8-bromodeoxyguanosine-3'-phosphoramidite was used for the preparation of the modified oligonucleotide d(TGG^{Br}GGT). The latter was assembled using the standard solid phase β -cyanoethylphosphoramidite chemistry. The oligomers were detached from the support and deprotected by treatment with concentrated aqueous ammonia at 55 °C for 12 h. The combined filtrates and washings were concentrated under reduced pressure, redissolved in H₂O, analyzed and purified by high-performance liquid chromatography (HPLC) on a Nucleogel SAX column (Macherey–Nagel, 1000-8/46); using buffer A: 20 mM KH₂PO₄/K₂HPO₄ aqueous solution (pH 7.0), containing 20% (v/v) CH₃CN; buffer B: 1 M KCl, 20 mM KH₂PO₄/K₂HPO₄ aqueous solution (pH 7.0), containing 20% (v/v) CH₃CN; a linear gradient from 0 to 100% B for 30 min and flow rate 1 ml/min were used. The fractions of the oligomer were collected and successively desalted by Sep-pak cartridges (C-18). The isolated oligomers proved to be > 98% pure NMR.

Synthesis of *N*-[5-([5-([3-amino-3-(*N*-methylcarbamoyl) ethyl]amino)carbonyl)-1-methyl-1*H*-pyrrol-3-yl]amino)carbonyl)-1-methyl-1*H*-pyrrol-3-yl]-4-(formylamino)-1-methyl-1*H*-pyrrole-2-carboxamide (2). A solution of 3-amino-*N*-methylpropanamide hydrochloride (123 mg, 1.2 mmol) in dry DMF (8

ml) was cooled to 5 °C and N, N'-diisopropylethylamine (206 ml, 1.2 mmol) was added. After 10 min, the acid **1** (412 mg, 1 mmol) and then EDCI (384 mg, 2 mmol) were added. The reaction mixture was stirred at room temperature for 18 hours, the solvent evaporated in vacuo and the crude residue purified by flash chromatography. The crude product purified by chromatography on a silica gel column, using DCM: MeOH (8:2) as eluent, furnished the compound **2** as a white solid (367 mg, 74% yield). Mp = 156-158 °C; IR (KBr): 3420, 3310, 1665, 1554, 1312 and 1212 cm⁻¹; ¹H-NMR (DMSO-d₆) δ: 2.42 (m, 2H), 2.77 (m, 3H), 3.34 (m, 2H), 3.78 (s, 3H), 3.82 (s, 3H), 3.83 (s, 3H), 6.65 (d, J = 1.6 Hz, 1H), 6.83 (d, J = 1.6 Hz, 1H), 6.82 (d, J = 1.6 Hz, 1H), 7.01 (d, J = 1.6 Hz, 1H), 7.17 (d, J = 1.6 Hz, 1H), 7.22 (d, J = 1.6 Hz, 1H), 7.89 (bs, 1H), 8.08 (s, 1H), 8.13 (s, 1H), 8.18 (s, 1H); 9.91 (s, 1H), 9.94 (s, 1H). FAB-MS (MALDI-TOF): 497.8 [M+1]⁺.

Synthesis of 4-Amino-N-[5-({[5-({[(3-amino-3-(N-methylcarbamoyl) ethyl] amino} carbonyl)-1-methyl-1H-pyrrol-3-yl]amino}carbonyl)-1-methyl-1H-pyrrol-3-yl]-1-methyl-1H-pyrrole-2-carboxamide (3**)**. To 5 ml of a solution obtained by the addition of 24 ml of ethanol (95%) to 1 ml of 36% HCl in water was added compound **2** (248 mg, 0.5 mmol), and the mixture stirred at room temperature for 24 hours. After this time, the reaction was filtered, the residue washed with dry ethyl ether (5 ml) and then dried on P₂O₅, to furnish compound **3** as brown solid, m.p. = 262-264 °C; ¹H-NMR (DMSO-d₆) δ: 2.41 (m, 2H), 2.56 (m, 3H), 3.24 (m, 2H), 3.78 (s, 3H), 3.80 (s, 3H), 3.81 (s, 3H), 6.75 (d, J = 1.6 Hz, 1H), 6.82 (d, J = 1.6 Hz, 1H), 6.84 (d, J = 1.6 Hz, 1H), 6.96 (d, J = 1.6 Hz, 1H), 7.04 (d, J = 1.6 Hz, 1H), 7.20 (d, J = 1.6

Hz, 1H), 7.88 (bs, 1H), 8.12 (s, 1H), 9.92 (s, 1H), 9.96 (s, 1H), 10.20 (bs, 3H). FAB-MS (MALDI-TOF): 469.4 $[M+1]^+$.

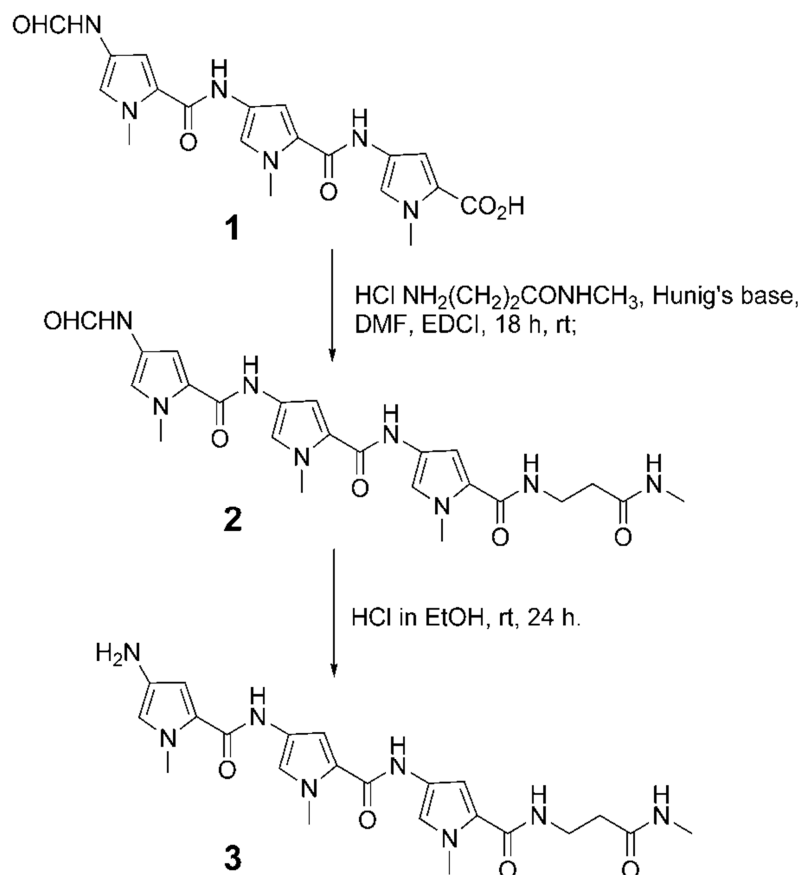


Figure 4.8: Chemical synthesis of derivative 3.

Nuclear magnetic resonance experiments. The quadruplex NMR sample was prepared at a concentration of 2 mM (8 mM single strand concentration), in 0.6 ml ($\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1) buffer solution having 10 mM KH_2PO_4 , 70 mM KCl, 0.2 mM EDTA, pH 7.0. For D_2O experiments, the H_2O was replaced with D_2O by drying down the sample, followed by lyophilization and redissolution in D_2O alone. NMR spectra were recorded with Varian ^{Unity}INOVA 700 MHz spectrometer. ^1H chemical shifts

were referenced relative to external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). 1D proton spectra of the sample in H₂O were recorded using pulsed-field gradient DPGSE^{5,6} for H₂O suppression. Phase-sensitive NOESY spectra¹⁰ were recorded with mixing times of 100 and 200 ms (T = 25 °C). Pulsed-field gradient DPGSE^{5,6} sequence was used for NOESY experiments in H₂O. TOCSY¹¹ and zTOCSY¹² spectra with mixing times of 100 ms were recorded with D₂O solution.

All experiments were recorded using STATES-TPPI¹³ procedure for quadrature detection. In all 2D experiments, the time domain data consisted of 2048 complex points in t₂ (16K in the zTOCSY) and 400–512 fids in t₁ dimension. The relaxation delay was kept at 3s for NOESY experiments used in the structure determination. A relaxation delay of 1.2 s was used for all other experiments. The NMR data were processed on a SGI Octane workstation using FELIX 98 software (Accelrys, San Diego, CA) and on iMAC running iNMR software (www.inmr.net).

Isothermal Titration Calorimetry. ITC measurements were conducted at 25 °C on a CSC 4200 Calorimeter from Calorimetry Science Corporation (Lindon, Utah). In these experiments, 10 µL aliquots of a 1250 µM ligand solution were injected into a sample cell containing 1300 µL of 40 µM quadruplex solution, with stirring at 297 rpm, for a total of 25 injections. The delay between injections was 400 seconds. Each ligand injection produced a heat burst curve. The areas under these heat burst curves were determined by integration to yield the associated injection heats. The heats of dilution were determined in parallel experiments by injecting a ligand solution of the same concentration in the same buffer. The heat of dilution was subtracted from the interaction heats prior to curve fitting. The

corrected heat values are plotted as a function of the molar ratio, to give the corresponding binding isotherms. The resulting isotherms were then fitted to a single set of identical sites model using the Bindwork program supplied with the instrument, to give the binding enthalpy (ΔH°), equilibrium binding constant (K_b), and stoichiometry (n). The remaining thermodynamic parameters, ΔG° and $T\Delta S^\circ$, were derived using the standard relationships $\Delta G^\circ = -RT \ln K_b$ and $T\Delta S^\circ = \Delta H^\circ - \Delta G^\circ$. The buffer conditions used for ITC measurements were the same of NMR experiments.

Structure calculations. Restrained Simulated Annealing (SA) calculations were performed using the AMBER 10.0 package¹⁴, with DNA described by the latest AMBER force field for nucleic acids (ff99bsc0)^{15,16} while compound **3** was parameterized using the General AMBER Force Field¹⁷. The Jaguar software package (Schrodinger) was used to calculate atomic charges using DFT with an HF/6-31G** basis set at the B3LYP level.

For the initial AMBER model, compound **3** was manually docked into two opposite grooves of the [d(TGGGGT)]₄ using as a reference structure the distamycin A/[d(TGGGGT)]₄ solution structure (PDB code 2JT7)³. In fact, given the presence of many head to tail ligand-ligand NOEs, as happened for distamycin A, also **3** is expected to bind the quadruplex structure in a dimeric antiparallel form. Each **3** dimer was then placed into the quadruplex groove consistently with what recorded in NMR experiments.

For annealing simulations the General Born solvation (igb = 2) with monovalent salt concentration corresponding to 0.1 M was used. The complex was

heated to 900 K in the first 5 ps, cooled to 100 K for the next 13 ps and then cooled to 0 K for the last 2 ps. The temperature of the system was maintained with a varying time constant: 0.4 ps during heating, 4 ps during cooling to 100 K, 1 ps for the final cooling stage, and then reduced from 0.1 to 0.05 for the last picosecond. The force constants for NOE constraints were increased from 3 to 30 kcal mol⁻¹ Å⁻² during the first 5 ps and then maintained constant for the rest of the simulation. These force constants were applied in the form of a parabolic, flat-well energy term where r is the model distance or torsion angle and k is the respective force constant.

$$E_{\text{constraint}} = k(r_2 - r)^2 \quad r_1 \leq r < r_2$$

$$E_{\text{constraint}} = 0 \quad r_2 \leq r \leq r_3$$

$$E_{\text{constraint}} = k(r_3 - r)^2 \quad r_3 \leq r < r_4$$

The values for r_1 and r_4 represent upper and lower distance bounds, defining the linear energetic penalty before and after the flat-well energy term. The applied distance NOE constraints were retrieved by cross-peak volume integrations performed with the program FELIX 98 (Accelrys, San Diego, USA), using the NOESY experiment collected at mixing time of 100 ms. The NOE volumes were then converted to distance restraints after they were calibrated using known fixed distances. The NOE restraints were generated with three distance classifications as follows: strong NOEs ($r_2 = 1.0$ Å; $r_3 = 3.5$ Å), medium NOEs ($r_2 = 3.0$ Å; $r_3 = 4.5$ Å) and weak NOEs ($r_2 = 4.0$ Å; $r_3 = 6.0$ Å). Distance constraints between base pairs, set at half the NOE distance constraint strength, were applied to maintain hydrogen bonding (as indicated by the observation of imino proton resonances). Hydrogen bonds constraints were used in the range of 1.7 Å (r_2)/ 2.3 Å (r_3). Planarity force

constraints of $200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ were applied to G bases throughout the simulations. In agreement with NMR data (see Results and Discussion section), the backbone torsion angles α , β , γ , δ and ϵ were restrained in the range $-150^\circ/-30^\circ$, $-230^\circ/-110^\circ$, $20^\circ/100^\circ$, $95^\circ/175^\circ$ and $-230^\circ/-110^\circ$, respectively⁷. Further, glycosidic torsion angles χ were fixed in the *anti* domain ($-155^\circ/-75^\circ$). Interactions within the system were calculated. An unrestrained energy minimization step completed the simulated annealing run. This simulated annealing/energy minimization procedure was repeated 100 times. Ten best models were selected based on the value of the overall potential energy and NMR restraint violations for further analysis. After charge neutralization by the addition of 20 K^+ ions, the complexes were solvated with 16758 water molecules in a TIP3P pre-equilibrated box¹⁸. Several equilibration steps were performed comprising minimization of the solvent molecules with the DNA and ligand fixed, minimization of the whole system, and slow heating to 300 K with weak positional restraints on DNA and drug atoms under constant-volume conditions. The following 5 ns production runs were applied in the NPT ensemble. The particle mesh Ewald method¹⁹ was used to evaluate the electrostatic interactions with a direct space sum cutoff of 10 Å. With the bond lengths involving hydrogen atoms kept fixed with the SHAKE algorithm, a time step of 2 fs was employed²⁰.

Related conformational substates populated during the molecular dynamics simulation were analyzed with the AMBERS's PTRAJ module²¹. For the 10 trajectories the average configuration was taken as a reference for subsequent mass-weighted rmsd calculations considering all atoms excluding terminal T bases.

The single snapshot of each trajectory with the lowest rmsd was taken as the representative dynamic structure of that simulation.

The final set of coordinates has been deposited in the Protein Data Bank (accession code: 2KVY).

References

- [1] Grehn L., Ragnarsson U. *J. Org. Chem.* **1981**, *46*, 3492-3497.
- [2] Juaristi E., Quintana D., Lamatsch B., Seebach D. *J. Org. Chem.* **1991**, *56*, 2553-2557.
- [3] Martino L., Virno A., Pagano B., Virgilio A., Di Micco S., Galeone A., Giancola C., Bifulco G., Mayol L., Randazzo A. *J Am Chem Soc.* **2007**, *129*, 16048-16056.
- [4] Pelton J. G., Wemmer D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393-1399.
- [5] Hwang T. L., Shaka A. J. *J. Magn. Res.* **1995**, *A112*, 275-279.
- [6] Dalvit C. *J. Biomol. NMR* **1998**, *11*, 437-444.
- [7] Kim S., Lin L., Reid B.R. *Biochemistry*, **1992**, *31*, 3564-3574.
- [8] Pagano B., Mattia C. A., Giancola C. *Int. J. Mol. Sci.* **2009**, *10*, 2935-2957.
- [9] Chaires J. B. *Arch. Biochem. Biophys.* **2006**, *453*, 26-31.
- [10] Jeener J., Meier B., Bachmann H. P., Ernst R.R. *J. Chem. Phys.* **1979**, *71*, 4546-4553.
- [11] Braunschweiler L., Ernst R. R. *J. Magn. Reson.* **1983**, *53*, 521-528.
- [12] Trippleton M.J., Keeler J. *Angew. Chem. Int. Ed.* **2003**, *42*, 3938-3941.
- [13] Marion D., Ikura M., Tschudin R., Bax A. *J. Magn. Reson.* **1989**, *85*, 393-399.
- [14] Case D. A. *et al.* AMBER; 10 ed.; University of California: San Francisco, CA, **2008**.
- [15] Wang J. M., Cieplak P., Kollman P. A. *J. Comput. Chem.* **2000**, *21*, 1049-1074.
- [16] Perez A., Marchan I., Svozil D., Sponer J., Cheatham T. E. 3rd, Laughton C. A., Orozco M. *Biophys. J.* **2007**, *92*, 3817-3829.
- [17] Wang J. M., Wolf R. M., Caldwell J. W., Kollman P. A., Case D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
- [18] Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W., Klein M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.
- [19] (a) Darden T., York D., Pedersen L. *J. Chem. Phys.* **1993**, *98*, 10089-10092. (b) Essmann U., Perera L., Berkowitz M. L., Darden T., Lee H., Pedersen L. G. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
- [20] van Gunsteren W. F., Berendsen H. J. C. *Mol. Phys.* **1977**, *34*, 1311-1327.

- [21] Shao J., Tanner S. W., Thompson N., Cheatham T. E.III. *J. Chem. Theory Comput.* **2007**, 3, 2312–2334.

5) Tandem application of virtual screening and NMR experiments in the discovery of brand new DNA quadruplex groove binders

Recently, our research group has unambiguously demonstrated a groove binding mode between distamycin A and the quadruplex [d(TGGGGT)]₄¹ (Section 2). So far, this compound remains the only comprehensively documented groove binder for DNA quadruplex. Classical intercalation, as observed in the duplex DNA structures, have not been demonstrated to date, most probably due to the presence of cations like K⁺ or Na⁺ in the very center cavity of the quadruplex structures, that prevents such a binding mode.

Generally, stacking interaction and groove-binding modes are characterized by very different specificities. Groove-binding recognition generally offers a higher extent of selectivity, since it can more easily recognize different DNA sequences and, in the case of quadruplex structures, it can also discriminate among several quadruplex topologies, taking advantage of their different groove widths. However, in order to find new quadruplex groove binding agents, a consideration must be made about the chemical nature of the quadruplex grooves. Interestingly, all known quadruplex structures are characterized by grooves that are chemically and conformationally very different from the minor groove of the duplex DNA. This means that searching for new quadruplex groove binding agents among duplex minor groove binders may not be the most successful strategy. For this reason, a number of research groups have recently focused their attention in finding alternative molecular scaffolds able to recognize the groove of the quadruplex^{2,3}. Here, we attempted to search for brand new molecular scaffolds able to interact

with the groove of quadruplex structures by means of a structure-based virtual screening (VS) approach⁴. Actually, whereas there are numerous studies using these screening methods for targeting proteins, only few VS campaigns have been undertaken targeting nucleic acids, and, to the best of our knowledge, none targeting the groove of the quadruplex. In a recent review, Trent and co-workers⁵ have outlined that the software Autodock optimally balances docking accuracy and ranking. Thus, as a starting point, this program was used in VS experiments aimed at targeting a very simple quadruplex, namely [d(TGGGGT)]₄ (PDB code 1S45)⁶. This quadruplex possesses a fourfold symmetry with all strands parallel to each other, which afford four grooves of identical medium width, and all nucleosides in an *anti* glycosidic conformation. Thus, the docking software Autodock4 (AD4) was used to dock a diversity set of the commercially available Life Chemicals database (6000 compounds). In order to avoid finding redundant information, a search area large enough to enclose only one of the four identical grooves was used (Figure 5.1). The VS results were sorted on the basis of their predicted binding free energies (ΔG_{AD4}) which ranged from -0.95 to -9.55 kcal/mol. Solutions with a predicted binding free energy greater than -6.0 kcal/mol and a cluster size lower than 10 out of 100 individuals were discarded. Based on these criteria only 137 individuals were retained for further considerations. The binding poses calculated for these compounds were then visually inspected to discard all the individuals which were not predicted to establish tight interactions with the groove of the quadruplex structure. More precisely, compounds that were not able to form H-bond with any of the guanine bases and/or to establish electrostatic interaction with the backbone phosphate groups were not considered. After this final step, 30 compounds

corresponding to 0.5% of the original Life Chemicals database were selected and purchased for further analysis.

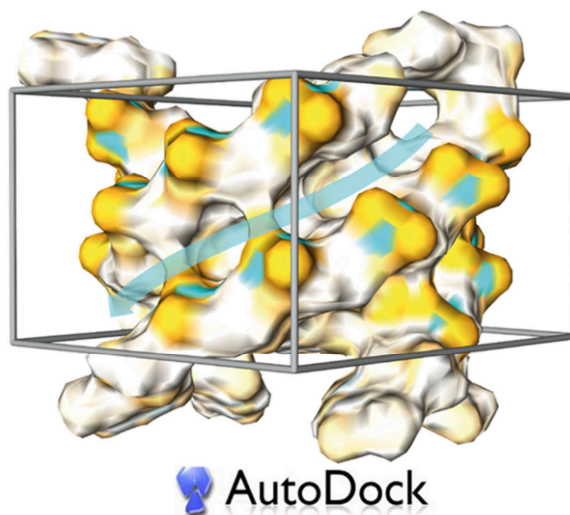


Figure 5.1: Groove of the quadruplex [d(TGGGGT)]₄ used as VS search area

The experimental test of the top computational “hits” for binding has been performed by NMR, which has the significant advantages that it can detect weak binders and readily identify the ligand binding sites, consequently verifying the binding specificity⁷. Thus, we prepared 30 identical DNA samples containing 6.8 mM of d(TGGGGT) (meaning 1.7 mM concentration of quadruplex), in 0.2 ml (H₂O/D₂O 9:1) buffer solution having 10 mM KH₂PO₄, 70 mM KCl, 0.2 mM EDTA, at pH 7.0. The first problem that we encountered in testing the selected compounds was solubility (most of them were not soluble in water). Nevertheless, all samples turned out to be soluble in DMSO. Therefore, we tested the stability of the quadruplex [d(TGGGGT)]₄ in buffers containing different % of DMSO. The quadruplex turned out to be perfectly structured even in buffer solutions containing 30% of DMSO. Therefore, we dissolved each equivalent of the selected compounds in 5 μ L of

DMSO, in such a way to get an overall percentage of DMSO at the end of the titration not higher than 15%.

The NMR titrations (700 MHz, T = 25 °C) were carried out monitoring resonance chemical shift changes of DNA, which were used to estimate whether a given compound is able to interact with the quadruplex and to determine the binding site.

In line with the expected results of a VS campaign, a number of false positive were found: i) 8 molecules do not significantly interact with the quadruplex, ii) 11 molecules caused shift of the signal of Ts (T1 and T6), iii) 3 molecules caused shift of the resonances belonging to external bases (T1 and G2, T6 and G5), most probably due to an end-stacking interaction.

Outstandingly, six molecules (**1-6**, Figure 5.2) were found to cause an appreciable shift, among the others, of the G3 and G4 resonances signals, thus suggesting, as expected, a groove binding interaction.

Particularly, compounds **1-5** provided NMR titration profiles very similar to each other, causing mainly drifting of the signals of G3, G4, G5 and T6, and so indicating that the recognition process involves mostly the 3' side of the grooves. This is a very interesting result, since it is surprisingly consistent with the mode of binding calculated by the VS. On the other hand, **6** seems to entirely span the grooves, perturbing more uniformly all the residues of the quadruplex. In addition, **6** causes a major change of the resonances of the quadruplex, suggesting a higher affinity. Interestingly, while **1-4** and **6** are positively charged as most of the already known groove binders, **5** does not possess any charge, so that the binding might be

more driven by the presence of a number of H-bond acceptor heteroatoms. The titration of all 6 molecules turned out to be virtually completed at 4 equivalents.

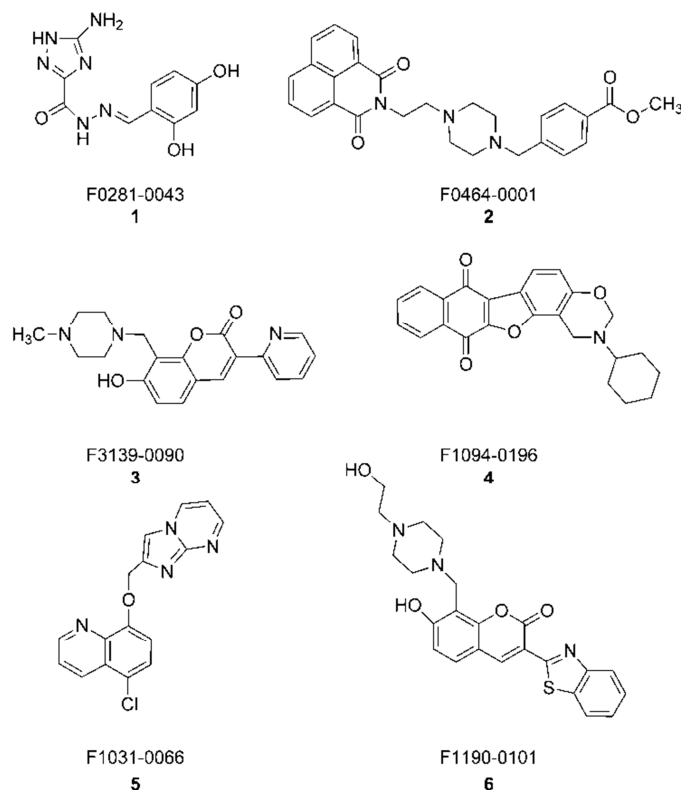


Figure 5.2: Structures of the newly identified groove binders. The Life Chemicals codes are reported in plain text. Numerals used in this paper are reported below in bold.

In summary, the application of VS calculations together with NMR experiments proved to be a successful strategy in the identification of new molecular chemotypes able to bind the grooves of DNA quadruplex structures. The structural diversity of these inhibitors has provided valuable alternative series for ongoing lead-optimization aimed at the identification of brand new pharmacological tools, endowed of a better affinity and pharmacokinetic profile, useful in the clarification of the mechanism, targeting, and therapeutic potential of

G-quadruplexes. In the future we plan to apply this lead discovery approach targeting different and more complex G-quadruplex structures.

References

- [1] Martino L., Virno A., Pagano B., Virgilio A., Di Micco S., Galeone A., Giancola C., Bifulco G., Mayol L., Randazzo A. *J. Am. Chem. Soc.* **2007**, 129, 15950-15956.
- [2] Dash J., Shirude P.S., Hsu S.D., Balasubramanian, S. *J. Am. Chem. Soc.* **2008**, 130, 16048-16056.
- [3] Li Q., Xiang J., Li X., Chen L., Xu X., Tang Y., Zhou Q., Li L., Zhang H., Sun H., Guan A., Yang Q., Yang S., Xu G. *Biochimie* **2009**, 91, 811-819.
- [4] Huey R., Morris G. M., Olson A. J., Goodsell D. S. *J. Comput. Chem.* **2007**, 28, 1145-1152.
- [5] Dailey M.M., Hait C., Holt P.A., Maguire J.M., Meier J.B., Miller M.C., Petraccone L., Trent J.O. *Exp. Mol. Pathol.* **2009**, 86,141-150
- [6] Caceres C., Wright G., Gouyette C., Parkinson G., Subirana J.A. *Nucleic Acids Res.* **2004**, 32, 1097-1102.
- [7] Pellecchia M., Bertini I., Cowburn D., Dalvit C., Giralt E., Jahnke W., James T.L., Homans S.W., Kesler H., Luchinat C., Meyer B., Oschkinat H., Peng J., Schwalbe H., Siegal G. *Nature Rev. Drug Discovery* **2008**, 7, 738-745.

6) A more detailed picture of the interactions between virtual screening-derived hits and the DNA G-quadruplex: NMR, molecular modelling and ITC studies

6.1 Introduction

The first evidence that the grooves of a quadruplex structure can be recognised by an organic molecule came approximately three years ago from our own NMR work revealing that, distamycin A can bind in a dimeric form to the two opposite grooves of the quadruplex¹. Whereas the end-stacker ligands are the major part of all known quadruplex binders and their number is growing each day, quadruplex groove binders represent a quite unexplored and valuable field. In fact, grooves in duplex and quadruplex DNA are chemically and conformationally different, and since groove dimensions differ according to the type of quadruplex, groove binders can in principle be selective for a particular quadruplex topology. Thus, with the aim of finding brand new molecular scaffolds able to interact with the groove, starting from the quadruplex structure found in complex with distamycin A, we recently performed an extensive structure based virtual screening (VS) campaign². As a result, six molecules (**1-6**, Figure 5.2, section 5) were found to be somehow groove binding agents. Herein, we have performed additional NMR titrations of the six compounds with modified quadruplexes, namely [d(TGG^{Br}GGT)]₄ and [d(TGGGG^{Br}T)]₄. These NMR spectroscopy experiments combined with molecular modelling studies, provided a more detailed picture of the interactions between each binder and the quadruplex DNA. Noteworthy, isothermal titration calorimetry (ITC) measurements on the above-mentioned compounds revealed that

2, **4**, and **6** despite their relatively small dimensions, are more effective in groove binding with respect to distamycin A, to the best of our knowledge, the most potent groove binder identified thus far.

6.2 Results and Discussion

NMR and molecular modeling studies. Using NMR, six molecules (**1-6**, Figure 5.2, section 5) were found to be potential groove binders: they cause an appreciable shift, amongst others, of the signals of G3, G4, G5 and T6 of the parallel quadruplex [d(TGGGGT)]₄, indicating that the recognition process involves mostly the 3' side of the grooves, as also computed by the Virtual Screening calculations².

In order to get further insights into the binding mode of compounds **1-6**, we have acquired a number of NOESY experiments of the complexes of the six compounds with the quadruplex [d(TGGGGT)]₄. Unfortunately, as no diagnostic NOE cross-peak could be retrieved for any complex, none of the three-dimensional structures at atomic level could be determined.

Thus, we decided to get a more detailed picture of the interactions from the NMR titration of modified quadruplexes and from molecular docking calculations. As a result, we designed and synthesized modified oligonucleotides, namely d(TGG^{Br}GGT), d(TGGG^{Br}GT) and d(TGGGG^{Br}T), where dG^{Br} is 8-bromo-2'-deoxyguanosine, potentially capable of forming quadruplex structures and possessing a bulky group (bromine) at different positions of the grooves (G3, G4 and G5), as this portion of the DNA should be involved in **1-6** recognition². These modified quadruplexes would then be titrated with the six molecules **1-6** and the

results of the NMR titration profile analyzed. The rationale behind these experiments derived from the assumption that if these compounds interacted with the quadruplex groove region, the presence of the bromine group should now prevent (or at least limit) the ligand/DNA interactions.

Prior to these experiments, we tested the capability of d(TGG^{Br}GGT), d(TGGG^{Br}GT) and d(TGGGG^{Br}T), to form a quadruplex structure. Their NMR samples were prepared at a concentration of 2 mM, in 0.6 ml (H₂O/D₂O 9:1) buffer solution having 10 mM KH₂PO₄, 70 mM KCl, 0.2 mM EDTA, pH 7.0. These samples were then annealed for 5-10 minutes at 80 °C and slowly cooled down to room temperature, then ¹H-NMR spectra were recorded by using DPGSE pulse sequence for H₂O suppression^{3,4}. The ¹H-NMR spectra (700 MHz, T = 25 °C) of d(TGG^{Br}GGT) and d(TGGGG^{Br}T) show the presence of four well defined singlets in the region 11-12 ppm, ascribable to imino protons involved in Hoogsteen hydrogen bonds of G-quartets, as well as the presence of five signals belonging to three guanine H8 and to two thymine H6 protons in the aromatic region. This indicates that a single well defined quadruplex species is present in solution for both molecules, consisting of four G-tetrads and possessing a fourfold symmetry with all strands parallel to each other. In such a quadruplex, each Br group faces the grooves and points outward. On the other hand, the 1D ¹H-NMR spectrum of d(TGGG^{Br}GT) shows the presence of a great number of signals in the regions of imino and aromatic protons, suggesting that d(TGGG^{Br}GT) is affected by structural heterogeneity, thus preventing its use in our experiments. Hence, only the quadruplexes [d(TGG^{Br}GGT)]₄ and [d(TGGGG^{Br}T)]₄ were titrated with compounds **1-6** and the titrations monitored by NMR. A comparison of resonances of protons of the uncomplexed quadruplex and the

complexed one has been performed. In particular, we report the $\Delta\delta$ values (chemical shifts of the complex minus free DNA) of aromatic, methyl and imino protons. As far as compound **1** is concerned, the titration with unmodified quadruplex [d(TGGGGT)]₄ led to a general shift of the monitored signals (white bars in Figure 6.1). On the other hand, the titration of [d(TGG^{Br}GGT)]₄ causes a slight shift (light-gray bars in Figure 6.1) of the residues at the 3' edge of the quadruplex, namely G5-H8, T6-H6/Me, and only of T1-H6/Me at 5' edge, whilst no appreciable shift can be measured for the other signals. This means that the bromine atom at the very centre of the groove did affect the binding of compound **1**, and that **1** can be confidently considered a groove binder. Furthermore, the titration of [d(TGGGG^{Br}T)]₄ led to a severe line broadening of all signals, making impossible to retrieve any information from the spectra. This NMR phenomenon can be interpreted assuming that the ligand is changing its binding pose on the NMR time scale. It is noteworthy that compounds **1-6** have been discovered from a previous Virtual Screening campaign and in that study Autodock4 program was used with a search area large enough to enclose only one of the four identical grooves to avoid redundant information². However, this couldn't cast out the possibility that one or more of the selected molecules can also be able to bind to other part of the target. Thus, herein, we present results of new docking calculations where the search area has been enlarged to comprise the entire surface of the quadruplex [d(TGGGGT)]₄ with the purpose of checking for the capability of the ligands to extend their binding out of the groove.

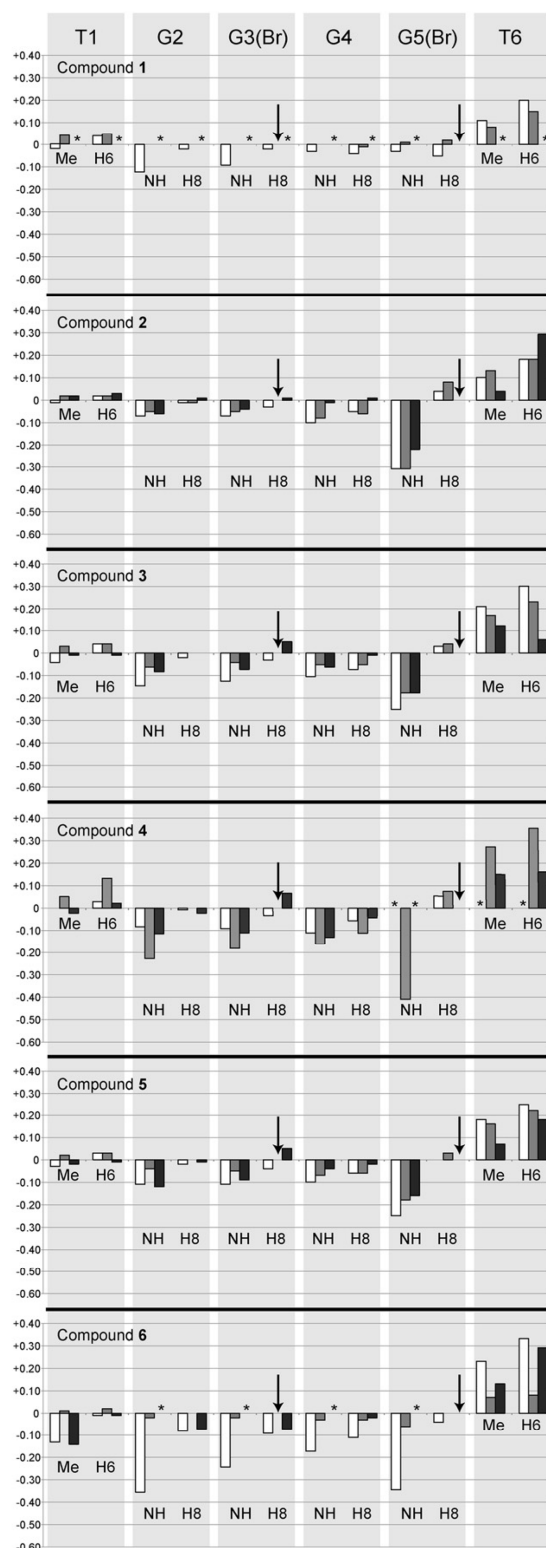


Figure 6.1: $\Delta\delta$ values of aromatic (H6/H8), methyl (Me) and imino protons (NH) for $[d(TGGGGT)]_4$ (white bars), $[d(TGG^{Br}GGT)]_4$ (light-gray bars) and $[d(TGGGG^{Br}T)]_4$ (dark-gray bars). Shadowed columns gather resonance variation belonging to the same residue. Asterisks indicate a severe line-broadening of the monitored signals. Arrows indicate the lack of bars due to the presence of bromine atoms.

Regarding **1**, docking calculations showed that all solutions in the lowest energy families ($\Delta G = -6,2$ kcal/mol), were found to be anchored to the 3' side of the groove in line with the above mentioned NMR data (**1** was predicted in a groove binding mode 75 out of 100 time. As shown in Figure 6.2a, the amine group of **1** engages a H-bond with the 4'O of the T6 nucleoside, while the ligand carbonyl group together with the *ortho*-hydroxyl group of the phenyl moiety forms H-bonds with G5 base. Three H-bonds have been also detected between **1** and the phosphate backbone. The above-described binding pose has been found 25 out of 100 times.

Surprisingly, compound **2** generates almost identical $\Delta\delta$ values in the titration with $[d(TGGGGT)]_4$ and $[d(TGG^{Br}GGT)]_4$ (white and light-gray bars, respectively, in Figure 6.1), having a general perturbation of the signals belonging to the 3' side of the groove. This can be interpreted by assuming that the bromine atom does not affect the binding of **2** suggesting that this compound should not preferentially bind the very centre of the groove. On the other hand, the titration of $[d(TGGGG^{Br}T)]_4$ (dark-gray bars in Figure 6.1) led to a marked shift of T6-H6/Me signals whereas only a slight shift of all the other monitored signals can be observed. This indicates that **2** prefers to interact with the 3' edge of the quadruplex.

A clearer picture of the ligand-quadruplex interaction could be obtained from docking calculations using again the unmodified quadruplex $[d(TGGGGT)]_4$ as target. The calculations highly converged toward one family of conformations ($\Delta G = -7.4$) (Figure 6.2b) in which the benzoisoquinolinedione ring stacks on the surface of the 3' terminal quartet, particularly between the T6 and G5 rings, and the

rest of the molecule inserts into the groove, where the main anchor point is represented by the protonated nitrogen of the piperazine ring, which establishes a charge reinforced H-bond with the phosphate backbone. This binding mode is in agreement with the new NMR titrations performed in this investigation indicating that **2** is actually characterized by a mixed binding mode, providing both stacking and groove binding interactions.

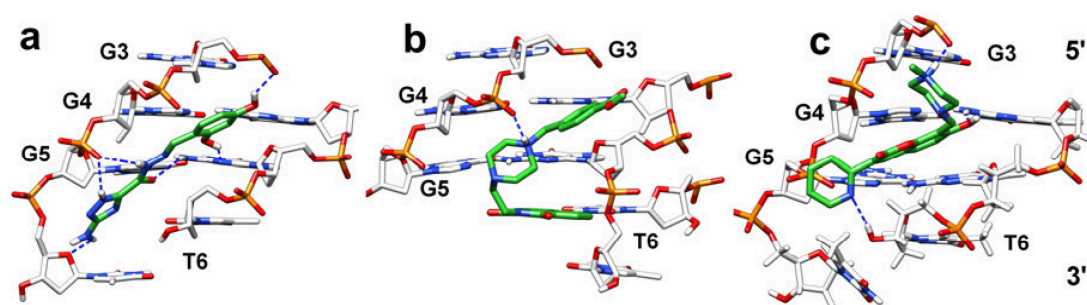


Figure 6.2: Binding poses calculated by AD4 for compounds **1** (a), **2** (b) and **3** (c) in the quadruplex structure. DNA backbone is represented as white stick bonds. Ligands are depicted as green sticks. H-bonds are represented as dashed blue lines.

Differently from **1** and **2**, titration of compound **3** with $[d(TGGGGT)]_4$, $[d(TGG^{Br}GGT)]_4$ and $[d(TGGGG^{Br}T)]_4$ led to a general shift of almost all signal (Figure 6.1). However, the resonances of $[d(TGGGGT)]_4$ (white bars) shifted more and those of $[d(TGG^{Br}GGT)]_4$ and $[d(TGGGG^{Br}T)]_4$ less (light- and dark-gray bars, respectively). This indicates that **3** does not possess a well defined binding mode, so that, when the groove is unavailable by the presence of the bromine atoms, **3** slides toward the 3' edge groove of the quadruplex; vice versa, in the case of $[d(TGGGG^{Br}T)]_4$, the molecule is able to interact only weakly with the available part of the groove. Docking calculations on $[d(TGGGGT)]_4$ suggest that the ligand binds the 3' end of the groove ($\Delta G = -8.2$) (**3** was predicted in a groove binding mode 87 out of 100 time).

As shown in Figure 6.2c, two H-bonds were found, the first between the pyridine nitrogen of **3** and the T6 3'-OH group and the second between the hydroxyl group of the chromenone core and the G3 NH₂. Hydrophobic contacts between the pyridine and the T6 rings were also detected together with a charge reinforced H bond established by the protonated nitrogen of the piperazine ring with the phosphate backbone. The above-described binding mode has been found by Autodock program 27 out of 100 times.

Interestingly, compound **4** displayed different behaviours for the three quadruplexes tested. With the quadruplex [d(TGG^{Br}GGT)]₄ (light-gray bars in Figure 6.1), the titration provided higher $\Delta\delta$ values for the residues at the edges of the quadruplex with the respect of unmodified [d(TGGGGT)]₄ (white bars in Figure 6.1). On the contrary, titration of [d(TGGGG^{Br}T)]₄ caused major shift for the residue sitting in the very centre of the groove (dark-gray bars). This means that compound **4** can interact with the grooves, and, accordingly, when the very centre of the groove is hindered, **4** binds the end sides of the grooves, but it is able to bind the very centre of the grooves when their 3' edges are unavailable. Docking calculations suggest two binding poses especially in line with NMR data (Figures 6.3a and 6.3b), where the molecule can either interact with 5' residues (binding mode A, $\Delta G = -8.2$) or with the 3' end (binding mode B, $\Delta G = -7.9$). Specifically, in A, the benzoquinone ring establishes a π - π interaction with the T1 ring, while an H-bond between the G2 NH₂ group and the carbonyl moiety is detected. In B, the two carbonyl moieties of the ligand are engaged in H-bonds with G4 NH₂ and T6 NH respectively. In both binding modes a charge reinforced H-bond is observed between the protonated nitrogen of **4** and the phosphate backbone. Based on this finding and on the shift of

NMR signals observed for all residues forming the groove, a sliding motion of **4** inside the groove can be proposed. Noteworthy, 100 out of 100 runs propose **4** as groove binder.

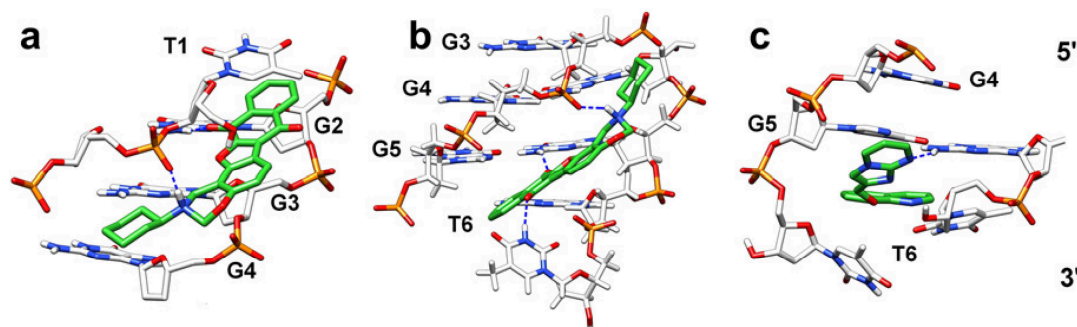


Figure 6.3: Two binding poses calculated by AD4 for compound **4** (a and b), and one for compound **5** (c) in the quadruplex structure. DNA backbone is represented as white stick bonds. Ligands are depicted as green sticks. H-bonds are represented as dashed blue lines.

As far as compound **5** is concerned, the titration of $[d(TGGGGT)]_4$ (Figure 6.1, white bars) clearly indicates that **5**, like the other selected compounds, prefers to bind the 3' edge of the groove. The titration of $[d(TGG^{Br}GGT)]_4$ displays significant lower $\Delta\delta$ values for the imino protons (Figure 6.1, light-gray bars), while the shifts of the signal of G4, G5 and T6 remain substantially unchanged. On the other hand, the titration of $[d(TGGGG^{Br}T)]_4$ indicates that **5** is still able to interact with the 3' edge of the quadruplex probably via end-stacking interaction. A mixed binding mode is therefore expected for compound **5**. In line with these data, docking calculations suggested that in **5** the quinolone ring is adapted on the top of the 3' end sandwiched between the T6 and G5 bases (Figure 6.3c), with the imidazopyrimidine moiety extending towards the groove incipience and H-bonding with the G5 NH_2 group ($\Delta G = -6.7$). The above-described binding mode has been proposed by the docking program 60 out of 100 times.

Finally, compound **6** displays peculiar titration behaviours. While the titration of the unmodified [d(TGGGGT)]₄ showed marked shifts of the G2, G3, G4 and G5 residues (Figure 6.1, white bars), the titration of [d(TGG^{Br}GGT)]₄ is instead characterized by very low $\Delta\delta$ values for all residues (light-gray bars). Furthermore, the titration of [d(TGGGG^{Br}T)]₄ mainly caused shifts of the residues G2, G3 and T6, along with a severe line broadening of the imino proton signals (dark-gray bars, Figure 6.1). This means that **6** does bind the grooves of the quadruplex, and, as in **4**, the uniform perturbation of all groove residues of the quadruplex [d(TGGGGT)]₄, with respect to the limited ligand size, suggests that a fast sliding motion of **6** inside the groove is also possible. Molecular docking calculations also revealed that two molecules of **6** could in principle simultaneously anchor the two opposite ends of the groove although the 5' end is preferred (42 times out of 100) (Figures 6.4a and 6.4b). Notably, in perfect line with the changes experimentally observed for G2 and G5 aromatic protons signals, the ligand alternatively forms H-bonds with G2 ($\Delta G = -7.5$) or G5 ($\Delta G = -6.1$) bases in the two reported binding modes (Figure 6.4).

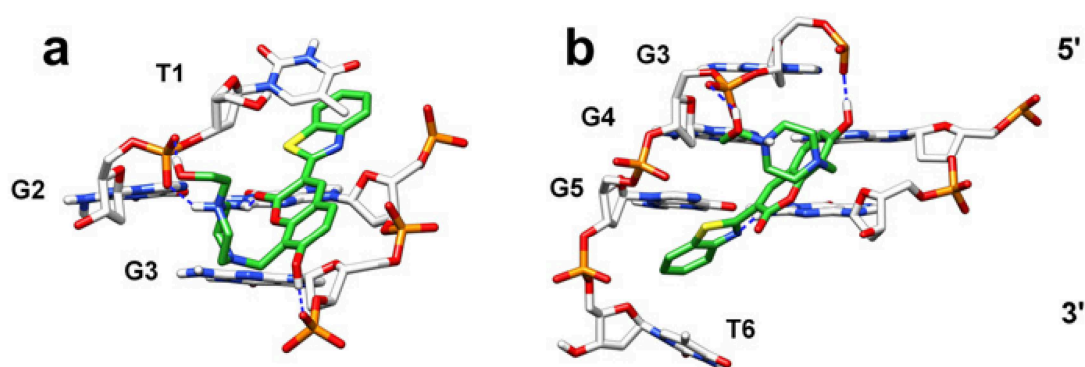


Figure 6.4: Two binding poses calculated by AD4 for compound **6**. DNA backbone is represented as white stick bonds. Ligands are depicted as green sticks. H-bonds are represented as dashed blue lines.

Finally, the two poses suggested by the docking program, in which a π - π interaction is formed with T1 or T6 ring, could somehow account for the T6-H6 proton and T1-methyl signals shift.

Isothermal Titration Calorimetry Measurements. In order to determine the binding affinity of the compounds **1-6** with the quadruplex [d(TGGGGT)]₄, ITC titration experiments were carried out⁵.

Previously, we characterised the interaction of distamycin A with the same quadruplex and revealed that four molecules of the drug were accommodated in two opposite grooves of the quadruplex; moreover, the thermodynamic signature of the binding event suggested that the entropic contribution was the one driving the complex formation¹. An analogous study however was not possible for the six new ligands identified with the virtual screening, because their poor solubility in water prevented us from performing canonical ITC experiments⁵, in which a solution of the quadruplex is titrated with a concentrated drug solution in the identical aqueous buffer. As an alternative approach to investigate the affinity of the new drugs for the DNA quadruplex, we carried out competition/displacement experiments, by analysing the ability of distamycin A to bind to the quadruplex in the presence of another compound^{6,7}. Despite the solubility concerns, mixtures of the quadruplex and each of the six new drugs were successfully prepared by solubilising the molecules in DMSO and diluting them in a phosphate solution containing the quadruplex (as described in the Methods); these complexes were then titrated with distamycin A, and the outcome followed by ITC.

In Figure 6.5 three representative ITC experiments are reported. First, the

interaction of distamycin A with [d(TGGGGT)]₄ was repeated in the new buffer used in this study in absence of any other ligand, clearly showing that in the new experimental conditions this binding event is identical to what was previously observed (Figure 6.5, Panels A and B and Table 6.1).

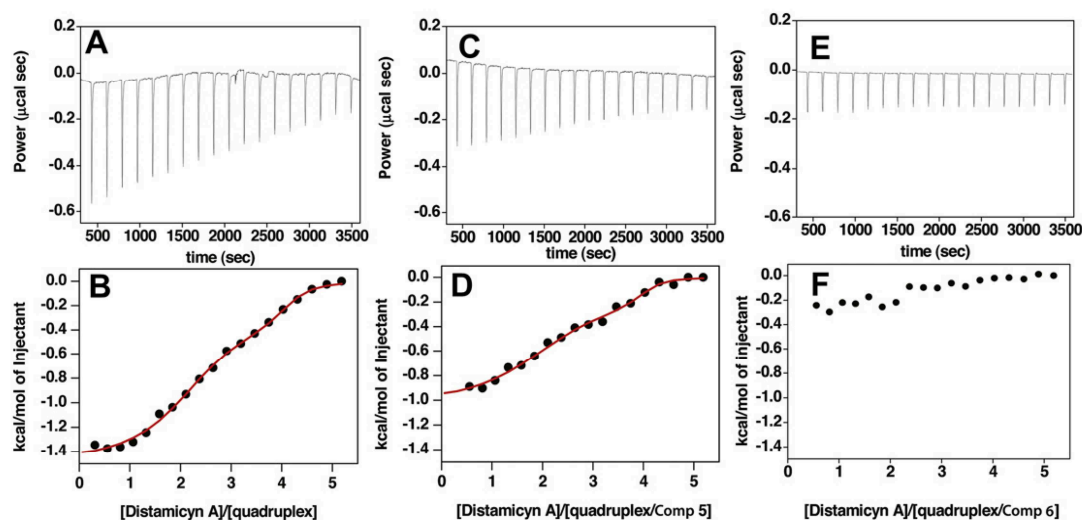


Figure 6.5: Calorimetric analysis of the interactions of the quadruplex [d(TGGGGT)]₄ and its complexes with the drugs **5** and **6**, with distamycin A. Raw titration data showing the thermal effect of injecting distamycin A into a calorimetric cell containing the quadruplex alone (A), quadruplex with **5** (C) and quadruplex with **6** (E). The normalised heat for the titrations, shown in B, D and F respectively, was obtained by integrating the raw data and subtracting the heat of the ligand dilution. The heat effect reveals the typical double step binding curve of the distamycin A interaction to the quadruplex in B and C, null interaction in F. The red lines in B and D represent the best fit derived by a non-linear least-squares procedure based on an independent binding sites model.

Compound in the calorimetric cell	<i>n</i>	<i>K_b</i> (M ⁻¹)	Δ <i>H</i> (kcal mol ⁻¹)	- <i>T</i> Δ <i>S</i> (kcal mol ⁻¹)	Δ <i>G</i> (kcal mol ⁻¹)
[d(TGGGGT)] ₄	1.8	(6.0 ± 0.2) × 10 ⁵	-0.3	-7.4	-7.7
	4.0	(7.0 ± 0.1) × 10 ⁶	-1.5	-7.7	-9.2
[d(TGGGGT)] ₄ /Compound 5	1.9	(3.0 ± 0.1) × 10 ⁵	-0.2	-7.1	-7.3
	4.0	(7.0 ± 0.1) × 10 ⁶	-1.0	-8.2	-9.2
[d(TGGGGT)] ₄ /Compound 3	1.7	(2.0 ± 0.1) × 10 ⁵	-0.6	-6.5	-7.1
	3.8	(7.0 ± 0.1) × 10 ⁶	-1.6	-7.6	-9.2
[d(TGGGGT)] ₄ /Compound 1	1.9	(3.0 ± 0.1) × 10 ⁵	-0.7	-6.6	-7.3
	3.8	(7.0 ± 0.1) × 10 ⁶	-1.6	-7.6	-9.2

Table 6.1: Thermodynamic parameters for the interaction of distamycin A with [d(TGGGGT)]₄ and its complexes formed with **1**, **3** and **5**. **2**, **4** and **6** are not reported since they abolish the binding of distamycin A

The binding occurs in two separate steps, the first one centred on a stoichiometry of 1:2 in which two molecules of distamycin A interact with the

quadruplex and then a second event in which two more molecules bind, leading to a final stoichiometry of 4 molecules of distamycin A per quadruplex. Panels C and E report two displacement experiments carried out titrating with distamycin A the complexes obtained with the drugs **5** and **6** respectively. The two experiments are quite different and indicative of two distinct phenomena. In the first case, a binding curve similar to the control experiment is obtained (compare Figures 6.5D and 6.5B), indicating that the presence of the drug **5** in the mixture did not have an effect on the association of distamycin A with the [d(TGGGGT)]₄. Distamycin A still binds in two distinct events, with stoichiometry and binding constants not significantly affected by the presence of **5** (Table 6.1). Conversely, the ITC titration of distamycin A into the quadruplex/**6** complex (Figures 6.5E and 6.5F) gives a completely different outcome. In this case, distamycin A appears no longer able to interact with the DNA, suggesting that **6** binds to the quadruplex tightly than distamycin A.

It is important to highlight that the competition/displacement experiments do not give any information about the stoichiometry of the complex formed between the quadruplex and the compound competing with the distamycin A. A summary of the results for all six compounds is reported in Table 6.1: **1** and **3** behave as **5**, whereas the presence of **2** or **4** or **6** abolishes the binding of distamycin A.

In conclusion the competition ITC experiments represent a possible strategy to evaluate whether an insoluble groove binder is stronger or weaker than a soluble drug that could be used as reference point. In our case, using the distamycin A as reference drug, we were able to group the six drugs identified with virtual screening

in two extreme classes. The first class contains compounds that bind rather weakly to the quadruplex (**1**, **3** and **5**), whereas in the second class the drugs that bind substantially stronger than distamycin A (**2**, **4** and **6**) are clustered.

6.3 Conclusions

A successful structure-based virtual screening (VS) campaign was recently undertaken by our research group and it resulted in the identification of six small molecules able to interact with the groove of the quadruplex [d(TGGGGT)]₄. Herein, we report NMR spectroscopy experiments which, combined with extensive molecular docking studies, allow for a more detailed picture of the interaction between each VS-derived binder and the quadruplex DNA. Noteworthy, isothermal titration calorimetry (ITC) measurements revealed that **2**, **4**, and **6**, despite their relatively small dimensions, bind substantially stronger than distamycin A, which is, to the best of our knowledge, the most potent groove binder identified so far. As it is widely accepted that specificity among the various DNA G-quadruplexes, that might be simultaneously present in the human genome, is a fundamental requirement for the quadruplex binder to become a drug, extensive binding tests towards DNA duplex and different DNA quadruplex topologies will be the next step in our research program. The future results combined with structural studies will provide a source of inspiration for the design of next generation of potent and selective quadruplex DNA drug-like binders.

6.4 Material and methods

Oligonucleotide synthesis. The oligonucleotide d(TGGGGT) was synthesized on a Millipore Cyclone Plus DNA synthesizer using solid phase β -cyanoethyl phosphoramidite chemistry at 15 μ mol scale. Commercially available 5'-DMT-aminoprotected-8-bromodeoxyguanosine-3'-phosphoramidite was used for the preparation of the modified oligonucleotides. The oligomers were detached from the support and deprotected by treatment with concentrated aqueous ammonia at 55 °C for 12 h. The combined filtrates and washings were concentrated under reduced pressure, redissolved in H₂O, analyzed and purified by high-performance liquid chromatography (HPLC) on a Nucleogel SAX column (Macherey–Nagel, 1000-8/46); using buffer A: 20 mM KH₂PO₄/K₂HPO₄ aqueous solution (pH 7.0), containing 20% (v/v) CH₃CN; buffer B: 1 M KCl, 20 mM KH₂PO₄/K₂HPO₄ aqueous solution (pH 7.0), containing 20% (v/v) CH₃CN; a linear gradient from 0 to 100% B for 30 min and flow rate 1 ml/min were used. The fractions of the oligomer were collected and successively desalted by Sep-pak cartridges (C-18). The isolated oligomers proved to be > 98% pure by NMR.

Nuclear magnetic resonance experiments. The quadruplex NMR samples were prepared at a concentration of 2 mM (8 mM single strand concentration), in 0.6 ml (H₂O/D₂O 9:1) buffer solution having 10 mM KH₂PO₄, 70 mM KCl, 0.2 mM EDTA, pH 7.0. NMR spectra were recorded with Varian ^{Unity}INOVA 700 MHz spectrometer. ¹H chemical shifts were referenced relative to external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). 1D proton spectra of the sample in H₂O

were recorded using pulsed-field gradient DPGSE^{3,4} for H₂O suppression. Phase-sensitive NOESY spectra⁸ were recorded with mixing times of 100 and 200 ms (T = 25 °C). Pulsed-field gradient DPGSE^{3,4} sequence was used for NOESY experiments in H₂O. NOESY experiments were recorded using STATES-TPPI⁹ procedure for quadrature detection. The time domain data consisted of 2048 complex points in t₂ and 400 fids in t₁ dimension. A relaxation delay of 1.2 s was used. The NMR data were processed on iMAC running iNMR software (www.inmr.net).

Molecular Modeling. The binding modes of compounds **1-6** were studied by means of docking experiments with the aid of Autodock4 (AD4)¹⁰ and using the [d(TGGGGT)]₄ G-Quadruplex DNA structure as deposited in the Protein Data Bank (PDB code 1S45) as macromolecules. The 3D structures of all the compounds were generated with the Maestro Build Panel¹¹. For the purpose of docking each molecule has been constructed in the protonation state suggested by the MarvinSketch 5.2.5.1 package (<http://www.chemaxon.com>) using a pH 7.0 accordingly with the NMR titrations. The target DNA structures were prepared through the Protein Preparation Wizard of the graphical user interface Maestro 9.0.211¹¹ and the OPLS-2001 force field. Water molecules were removed, hydrogen atoms were added and minimization was performed until the RMSD of all heavy atoms was within 0.3 Å of the crystallographically determined positions. Then the constructed compounds and DNA structures were converted to AD4 format files using ADT generating automatically all other atom values. In order to allow the ligands to explore all the possible search space, the docking area has been centred on the mass centre of the quadruplex structure and defined by a box large enough

to comprise the entire macromolecule. Accordingly, grids points of $84 \times 84 \times 84$ with 0.375 Å spacing were calculated around the docking area for all the ligands atom types using AutoGrid4. 100 separate docking calculations were performed for each binder. Each docking calculation consisted of 25×10^6 energy evaluations using the Lamarckian genetic algorithm local search (GALS) method. A low-frequency local search according to the method of Solis and Wets was applied to docking trials to ensure that the final solution represents a local minimum. Each docking run was performed with a population size of 150, and 300 rounds of Solis and Wets local search were applied with a probability of 0.06. A mutation rate of 0.02 and a crossover rate of 0.8 were used to generate new docking trials for subsequent generations. The docking results from each of the 100 calculations were clustered on the basis of root-mean square deviation (rmsd 2.0 Å) between the Cartesian coordinates of the ligand atoms and were ranked on the basis of the free energy of binding. All docking solutions were analyzed for the coherency with NMR data and for each compound, the lowest energy solution more in line with the experimental data was further considered and subjected to energy minimization through the OPLS-2001 force field. All Figures were rendered using Chimera software package¹².

Isothermal Titration Calorimetry. The [d(TGGGGT)]₄ stock solution was prepared by dissolving the lyophilised compound in 10 mM phosphate buffer with 70 mM KCl, 0.2 mM EDTA, pH 7. The solution was annealed by heating at 95 °C for 5 min and slowly cooling to room temperature. The concentration of the dissolved oligonucleotide was evaluated by UV measurement at 95 °C, using as molar extinction coefficient the value calculated by the nearest-neighbour model¹³ for the

sequence d(TGGGGT). Stock solutions of the six drugs were prepared by solubilising weighted amounts in DMSO to a final concentration of 8 mM. The complexes between the quadruplex and the drugs were prepared diluting the drug stock solution into the quadruplex solution to get a final DNA:drug molar ratio of 1:4 and a final DMSO concentration of 7%. Distamycin A was solubilised in the same buffer used for the oligonucleotide containing 7% of DMSO.

The titrations were carried out in 10 mM phosphate buffer, 70 mM, KCl, 0.2 mM EDTA, 7% DMSO, pH 7, at 293 K, using a high-sensitivity ITC-200 microcalorimeter from Microcal (GE Healthcare). In each experiment, volumes of 2 μ L of a 1.2 mM distamycin A solution were added into a 50 μ M solution of quadruplex-DNA complex, using a computer-controlled 40- μ L microsyringe, with a spacing of 180 s between each injection. Each titration was corrected for heat of dilution by subtracting the measured enthalpies of the injections following saturation. Integrated heat data obtained for the titrations were fitted using a nonlinear least-squares minimisation algorithm to a theoretical titration curve, using the MicroCal-Origin 7.0 software package from which the binding parameters ΔH° (reaction enthalpy change in $\text{kcal}\cdot\text{mol}^{-1}$), K_b (binding constant in M^{-1}), and n (stoichiometry) were derived. The entropic contribution was calculated using the relationships $\Delta G^\circ = -RT\cdot\ln K_b$ (R 1.987 $\text{cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$, T 293 K) and $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$.

References

- [1] a) Randazzo A., Galeone A., Mayol L. *Chem. Commun.* **2001**, 11, 1030-1031., b) Randazzo A., Galeone A., Esposito V., Varra M., Mayol L. *Nucleosides, Nucleotides & Nucleic Acids* **2002**, 21 535 - 545; c) Martino L., Virno A., Pagano B., Virgilio A., Di Micco S., Galeone A., Giancola C., Bifulco G., Mayol L., Randazzo A. *J. Am. Chem. Soc.* **2007**, 129, 15950-15956; d) Cosconati S., Marinelli L., Trotta R., Virno A., De Tito S., Romagnoli R., Pagano B., Limongelli V., Giancola C., Baraldi P. G., Mayol L., Novellino E., Randazzo A. *J. Am. Chem. Soc.* **2010** 132(18), 6425-6433; e) Pagano B., Virno A., Mattia C. A., Mayol L., Randazzo A., Giancola C. *Biochimie*, **2008**, 90(8), 1224-1232.
- [2] Cosconati S., Marinelli L., Trotta R., Virno A., Mayol L., Novellino E., Olson A.J., Randazzo A. *J. Am. Chem. Soc.* **2009**, 131, 16336-16337.
- [3] Hwang T. L., Shaka A. J. *J. Magn. Res.* **1995**, A112, 275-279.
- [4] Dalvit C. J. *Biomol. NMR*, **1998**, 11 437-444.
- [5] Pagano B., Mattia C. A., Giancola C. *Int. J. Mol. Sci.* **2009**, 10, 2935-2957.
- [6] Sigurskjold S.W., *Exact. Anal. Biochem.* **2000**, 277, 260-266.
- [7] Zhang Y.L., Zhang Z.Y. *Anal. Biochem.* **1998**, 261, 139-148.
- [8] Jeener J., Meier B., Bachmann H. P., Ernst R. R. *J. Chem. Phys.* **1979**, 71, 4546-4553.
- [9] Marion D., Ikura M., Tschudin R., Bax A. *J. Magn. Reson.* **1989**, 85, 393-399.
- [10] a) Huey R., Morris G.M., Olson A. J., Goodsell D. S. *J. Comput. Chem.* **2007**, 28, 1145-52; b) Cosconati S., Hong J. A., Novellino E., Carroll K. S., Goodsell D. S., Olson A. J. *J. Med. Chem.* **2008**, 51, 6627-6630.
- [11] Maestro, version 9.0.211, Schrodinger, L.L.C., New York, NY, **2009**.
- [12] Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Greenblatt D. M., Meng E. C., Ferrin T. E. *J. Comp. Chem.* **2004**, 25(13), 1605-1612.
- [13] Cantor C. R., Warshaw R. R., Shapiro H. *Biopolymers* **1970**, 9, 1059-1077.

7) Shooting for selective drug-like G-quadruplex binders: evidence for telomeric DNA damage and tumor cell death.

7.1 Introduction

Recently, our pursuit of new G-quadruplex ligands succeeded in the discovery of six lead-like chemotypes which were proven to effectively interact with the [d(TGGGGT)]₄ G-quadruplex structure¹. Among them, compound **6** renamed **1a** in the current work (Table 7.1), appeared to be the most promising hit. So far, ligands that selectively recognize G-quadruplex grooves are few², even though quadruplex groove recognition is likely to provide much more quadruplex-selective ligands. Interestingly, **1a** was proven to span the entire quadruplex grooves and was demonstrated to interact more tightly than distamycin A³, which was described as the most affine G-quadruplex groove binder⁴. More general considerations need to be done on the drug-like properties of the **1a** coumarin core. This is a naturally occurring structure (mainly in plants) which is present in a plethora of compounds endowed with different biological activities (anti-HIV, CNS-active, anti-coagulant, anti-inflammatory, antitumor)⁵. Thanks to the number of documented biological activities as well as its amenability to combinatorial chemistry, the coumarin scaffold represents a well-known example of privileged structure⁵.

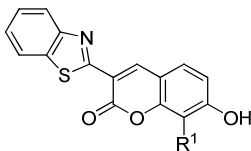
In the present study, the promising derivative **1a** was used as a seed for searching similar entities in several commercially available databases and NMR experiments allowed to identify a small focused library of structural analogues with G-quadruplex binding properties. By a back and forth approach, the structural features responsible for G-quadruplex groove recognition were delineated, while

isothermal titration calorimetry (ITC) measurements allowed for the identification of chemotypes featuring a tighter binding than distamycin A. Differently from distamycin A, the best binders were also proved to be G-quadruplex selective over duplex. These results propelled the biological characterization of the new ligands demonstrating their ability to induce selective DNA damage at telomeric level and induction of apoptosis and senescence on tumor cells.

7.2 Results and Discussion

Compound selection and G-quadruplex NMR binding assays. Given the relative synthetic accessibility of coumarin compounds we first decided to search in commercial molecular databases if analogues of compound **1a** were available. This approach has the advantage to rapidly provide a library of structural analogues of the lead compound being also highly economically efficient. Therefore, the simplistic pairwise Tanimoto similarity score was computed between **1a** and the compounds present in the full ZINC database collection of purchasable compounds (ca. 7 million compounds) using a 70% similarity threshold. This resulted in 272 compounds that were visually analyzed revealing that, as expected, coumarin derivatives were selected (Table 7.1) as well as several analogues featuring the regioisomeric chromone scaffold (Table 7.2). With the aim of exploring the influence of the pendant amine moiety on the quadruplex binding, five of the available coumarin derivatives deriving from the filtered database were first purchased (**1b-f**). These were all tested for their ability to interact with the DNA quadruplex through Nuclear Magnetic Resonance (NMR) spectroscopy, that is a

valuable tool to study molecular interactions in solution⁶. A number of DNA quadruplex folding topologies is known and the kind of folding is strongly affected by the sequence, buffer and molecular crowding conditions. This is particularly true for the human telomeric sequence. In fact, several structures have been reported having different strand orientations and loop distributions. Unfortunately, the topology of quadruplex structure adopted by the human telomeric sequence *in vivo* is unknown, so that, in analogy to what was done in our previous investigations^{1,4}, we have decided to use the highly symmetric tetramolecular quadruplex [d(TGGGGT)]₄. In particular, we employed the chemical shift perturbation method⁷ to detect interaction between a molecular candidate and the G-quadruplex structure. The signals that can be most easily monitored are the imino and aromatic protons of all bases and the methyl protons of thymines as they all are in the less crowded region of the NMR spectrum of [d(TGGGGT)]₄.



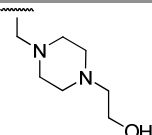
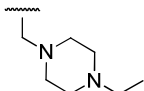
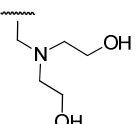
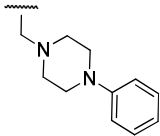
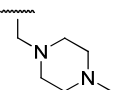
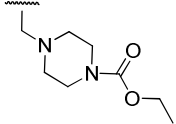
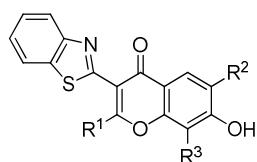
Cpd	R ¹	Cpd	R ¹
1a		1d	
1b		1e	
1c		1f	

Table 7.1: Structures of the selected coumarin derivatives



Cpd	R ¹	R ²	R ³	Cpd	R ¹	R ²	R ³
2a	-H	-H	-H	2k	-H	-H	
2b	-H	-H		2l	-NH ₂	-H	
2c	-H	-H		2m	-NH ₂	-H	
2d	-H	-H		2n	-NH ₂	-H	
2e	-H	-H		2o	-NH ₂	-H	
2f	-H	-H		2p	-NH ₂	-H	
2g	-H	-H		2q	-H	-CH ₂ CH ₃	
2h	-H	-H		2r	-H	-CH ₂ CH ₃	
2i	-H	-H		2s	-CH ₃	-H	
2j	-H	-H		2t	-CH ₃	-H	

Table 7.2: Structures of the selected chromone derivatives

Since we were mainly interested in potential groove binder analogues, in our inspection we considered quadruplex ligands all the compounds providing a shift of the G3 and G4 signals of at least 0.05 ppm. Following this criterion, among the five

tested coumarins, compounds **1b-d** demonstrated appreciable G-quadruplex binding while **1e** and **1f** displayed no DNA interaction. This indicates that the presence a H-bond donor (**1b**) and/or a positively charged group at a distal position (**1c-d**) with respect to the coumarin scaffold is critical for DNA binding.

More structural variability was present when considering the regioisomeric chromone derivatives (**2a-k**, Table 7.2). This moiety is also regarded as a privileged structure being present in different biologically active compounds and prone to combinatorial synthesis⁸. In this case, we first tested the G-quadruplex binding properties of **2a** through NMR titration. This experiment did not indicate appreciable shift of any signal of [d(TGGGGT)]₄ suggesting that the chromone scaffold is unable to provide efficient binding for the quadruplex. On the other hand, the positively charged group in position 8 (R³ substituent) should provide this moiety with a critical interaction point with the quadruplex (most probably, with the phosphate backbone atoms), as confirmed by the presence of detectable interactions of compound **2b** and **2c** with the target. Furthermore, bulky tertiary amines are not tolerated (**2d**), while cyclic amines (**2e-k**) are able to cause substantial shift of the G3, G4, G5 and T6 signals. In order to probe the influence of a substituent in position 2 on the chromone scaffold (R¹ in Table 7.2), **2l-2p**, featuring H-bond donor groups, were also tested. Interestingly, while **2l-2o** still provide appreciable quadruplex binding, **2p** does not interact with the DNA, suggesting that bulkier amines are detrimental for the binding regardless the presence of a H-bond donor in position 2. Indeed, the latter position seems to directly influence the G-quadruplex recognition considering that the simple substitution with a methyl group (**2s-t**) abolishes the ligand binding as

demonstrated by NMR experiments. The same holds true for position 6 (R^2) of the chromone ring that when substituted with a ethyl chain results in compounds (**2q-r**), which are incapable of $[d(TGGGGT)]_4$ binding.

Isothermal titration calorimetry (ITC) experiments. For compounds demonstrating appreciable interaction with the DNA G-quadruplex (**1a-b**, **2b-c**, **2e-o**) displacement isothermal titration calorimetry (ITC) experiments⁹ were performed to investigate their quadruplex binding affinity. Since it is well known that multiple conformations of the human telomeric sequences may coexist in solution and that such polymorphism could invalidate the data obtained by ITC measurements, we have decide to use, in analogy to NMR, the $[d(TGGGGT)]_4$ quadruplex.

Unfortunately, efforts to obtain direct thermodynamic information from canonical ITC experiments failed due to solubility issues regarding the ligands at the rather high concentrations required for such measurements¹⁰. Displacement experiments were then effectively carried out by analyzing the binding of distamycin A to the G-quadruplex previously saturated with each ligand (see Experimental Section). Although the displacement ITC experiments don't allow to directly measuring the thermodynamic parameters involved in the binding processes, this strategy represents a valid approach to evaluate the efficiency of a G-quadruplex (or duplex) binder compared to distamycin A. Indeed, when the DNA-ligand complex is formed, the ligand will inhibit the binding of distamycin A if its affinity for the G-quadruplex is higher than the latter; conversely, it will be displaced by a stronger binder.

Figures 7.1a and 7.1b show two representative ITC displacement experiments carried out by titrating with distamycin A the $[d(TGGGGT)]_4$ saturated with compounds **2o** and **2l**, respectively.

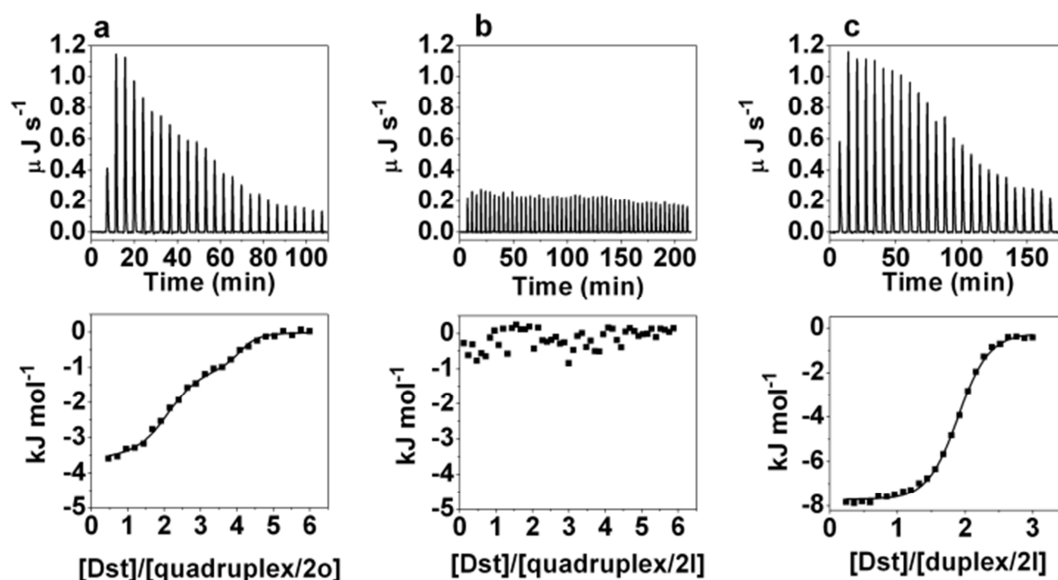


Figure 7.1: Raw ITC data (top panels) and integrated heat (bottom panels) for titration of $[d(TGGGGT)]_4/2o$ (a), $[d(TGGGGT)]_4/2l$ (b), and $d(CGCGAATTCGCG)_2/2l$ (c) mixtures with distamycin A (Dst) at 25 °C. The integrated heat for the titrations (squares) was obtained by integrating the raw data and subtracting the heat of the ligand dilution. The lines represent nonlinear least squares fit of the data to the appropriate binding model.

The results obtained in the two experiments clearly show two different phenomena. In the first case, the ITC profile for the titration of ligand-saturated G-quadruplex is essentially identical to the one obtained for the binding of distamycin A to the G-quadruplex alone³, thus indicating that the presence of **2o** does not affect significantly the interaction. On the contrary, the titration of $[d(TGGGGT)]_4/2l$ mixture with distamycin A (Figure 7.1b) gives completely different results. In this case, ITC data show constant heat release at each injection of distamycin A, only due to ligand dilution, proving that it is no longer able to interact with the G-quadruplex.

Overall, ITC experiments showed that four out of fifteen compounds (**1a**, **2c**, **2l** and **2m**) have the ability to bind the G-quadruplex more tightly than distamycin A, inhibiting its interaction. To evaluate a possible selectivity of the best ligands for G-quadruplex over duplex, we performed identical ITC displacement experiments by titrating d(CGCGAATTCGCG)₂ duplex/ligand mixtures with distamycin A. This self-complementary DNA dodecamer was chosen because it contains the central AATT core, considered being one of the specific binding sites for distamycin A¹¹.

Figure 7.1c shows an example of raw ITC data and binding isotherm for the titration of d(CGCGAATTCGCG)₂/**2l** mixture with distamycin A. As shown, at each injection of distamycin A solution, less and less heat release was measured until constant values were obtained, implying a saturable process. The binding isotherm shows a typical sigmoidal binding curve and clearly suggests that the presence of the compound in the mixture has no effect on the interaction of distamycin A with the duplex. Similar results were obtained for **2c** and **2m**, suggesting G-quadruplex selectivity of these three compounds. On the other hand, **1a** showed to affect distamycin A-duplex interaction, thus suggesting a poor selectivity.

The newly identified G-quadruplex ligands induce DNA damage and cell-cycle arrest. These encouraging results propelled the full biological characterization of the new ligands to investigate the ability of the new ligands to cause telomere uncapping (Figure 7.2). To this aim, a two-steps analysis was performed to establish, in a first one, if the compounds were able to induce DNA damage and, in a second one, if the DNA damage was localized to the telomeres.

By using human transformed BJ fibroblasts (BJ-HELT), we found that, differently from distamycin A, all the ligands were able to induce DNA damage (at least at the higher drug dose; Figure 7.2a) and deconvolution microscopy analysis showed that some of the damaged foci co-localized with TRF1, a good marker for interphase telomeres¹² forming the so-called TIFs (Telomere-dysfunction Induced Foci)¹³ (Figure 7.2d). Of note, quantitative analysis identified compounds **2c**, **2l** and **2m** as the most potent in inducing telomere damage: the percentage of cells with more than four γ H2AX/TRF1 colocalizations reached about 50% (Figure 7.2b), with a mean of about eight TIFs per nucleus (Figure 7.2c). Such a biological evaluation also confirmed that the sole chromone structure (**2a**) did not prove to cause substantial DNA damage.

Interestingly, the whole data are in perfect consonance with ITC experiments that indicated **2c**, **2l** and **2m** are strong selective G-quadruplex binders. A further correlation can also be done for compound **1a** that, by promiscuously binding the DNA (see ITC data), is able to induce high levels of DNA damage but few TIFs.

The above results raised the interesting possibility that telomere damages induced by the ligands in transformed fibroblasts may rapidly and efficiently promote growth inhibition in tumor cells. Treatment of Hela cells with one of the most promising selected ligands (**2l**), triggered a dose-dependent inhibition of cell survival, (Figure 7.3a) associated with an early accumulation of cells in the G₂/M phase of the cell cycle, and at 96 hrs. of treatment a fraction of cell population resided to the sub-G₁ compartment, indicative of apoptosis (Figure 7.2b).

Apoptosis induction triggered by **2l** has been confirmed by Annexin staining (Figure 7.3c; at 96 hrs. of treatment about 30% of cells are annexin V-positive/PI

negative) and it was also accompanied by the induction of a senescence phenotype: large cell size, vacuolated cytoplasm and β -galactosidase activity (Figure 7.3d).

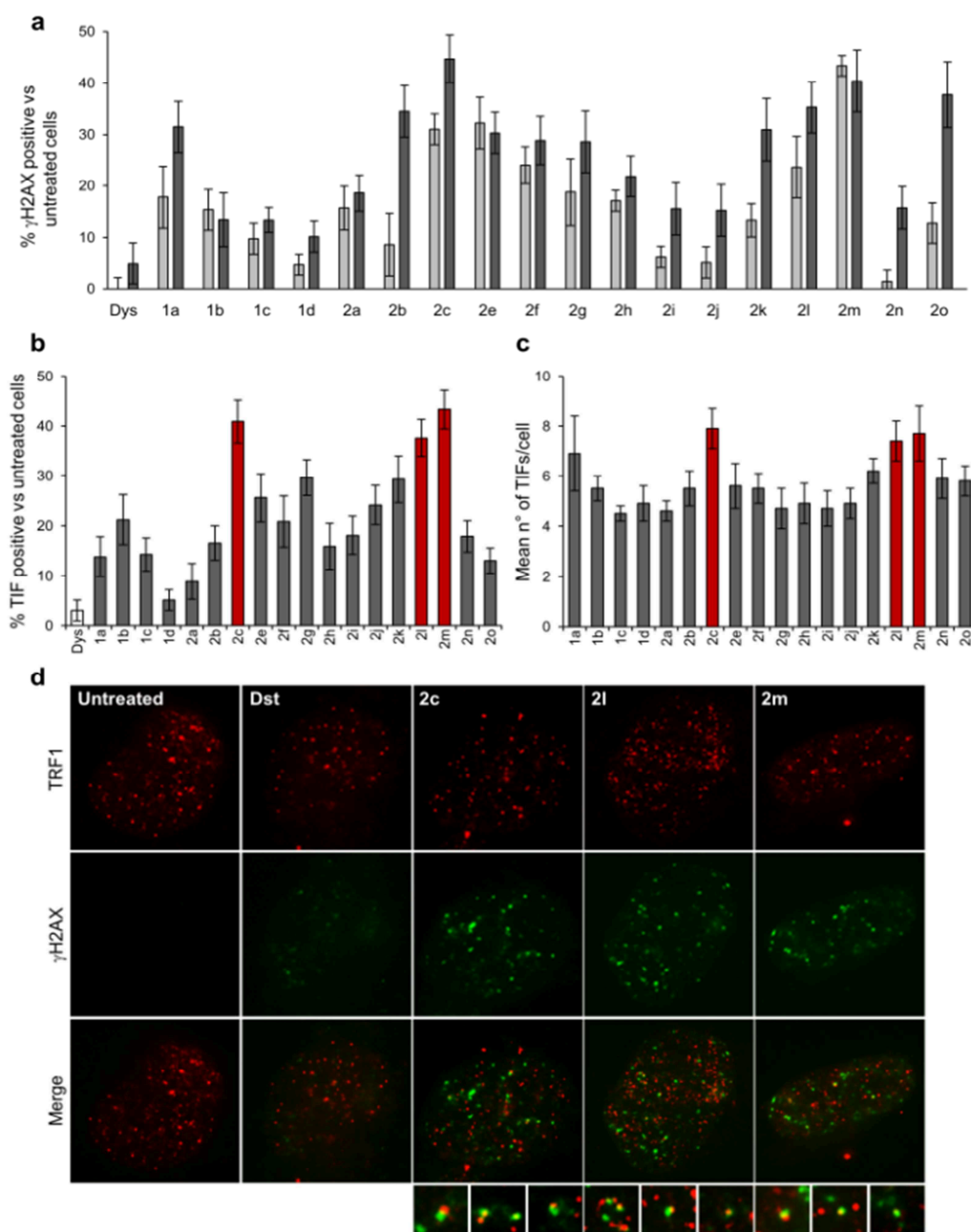


Figure 7.2: DNA damage activation at telomeres. BJ-EHLT fibroblasts were treated for 24 hrs. with distamycin A and the indicated ligands at the doses 0.1 (light-grey bars) and 0.5 μ M (dark-grey bars). Cells were processed for immunofluorescence (IF) using antibodies against γ H2AX and TRF1 to mark DNA damage and telomeres respectively. Percentages of γ H2AX- (a) and TIF-positive (b) treated vs untreated cells are reported in the histograms. (c) Mean number of TIFs in the indicated samples. Cells with four or more γ H2AX /TRF1 foci were scored as TIF positive. The red bars highlight the most effective ligands. Error bars indicate the standard deviation. (d) Representative images of IF of untreated and distamycin A, **2c**, **2l** and **2m**-treated BJ-EHLT cells. Enlarged views of TIFs are reported below the merged images. The images were acquired with a Leica Deconvolution microscope (magnification 100x).

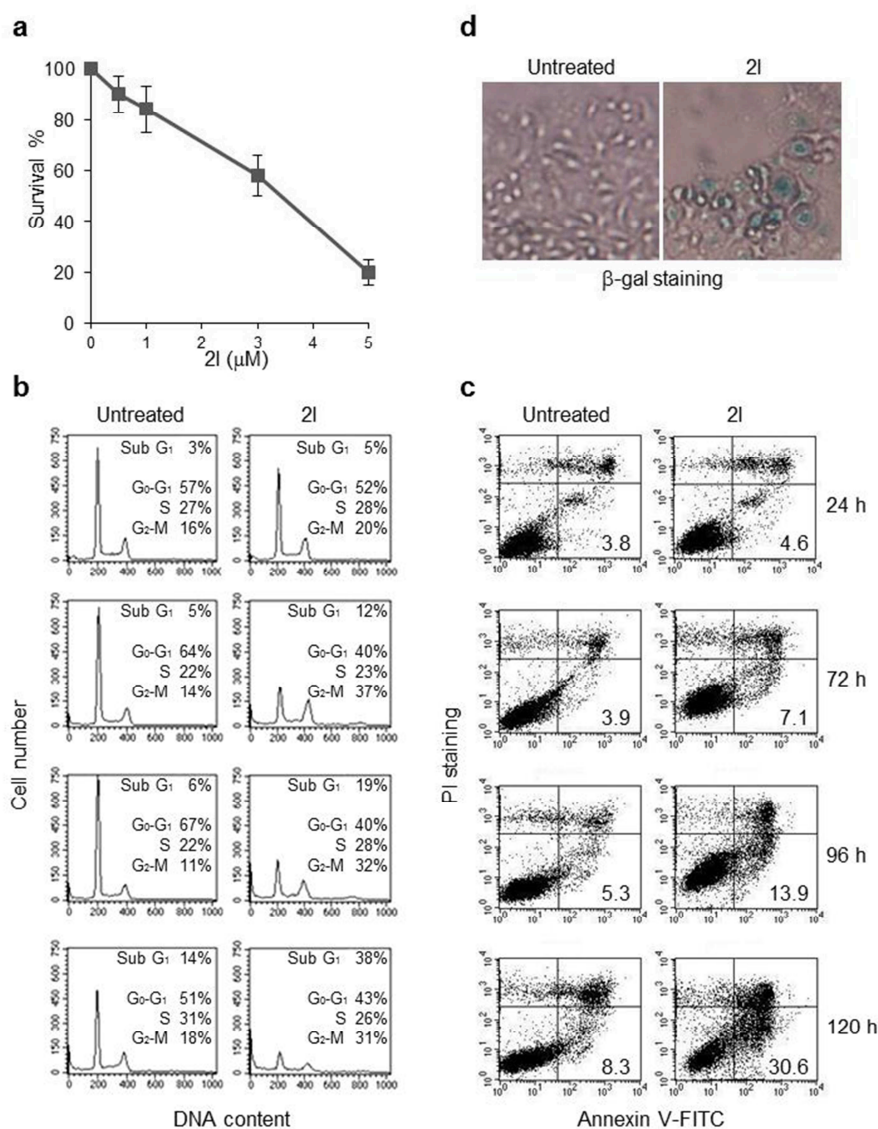


Figure 7.3: Biological effects of **2I** ligand. (a) Survival curve of the HeLa cells exposed to different doses of **2I** ranging from 0.5 to 5 μM . (b) Cell cycle analysis and (c) apoptosis evaluation of HeLa cells processed at the indicated times after exposure with 3.5 $\mu\text{M} **2I**. (c) Bi-parametric dot plots showing PI vs annexin V staining in the indicated samples. (d) SA-β-gal staining of HeLa cells untreated and treated with 5 $\mu\text{M} **2I** for 5 days.$$

The newly identified G-quadruplex ligands feature enhanced drug-likeness.

In our inspection, the most interesting ligands (**1a**, **2c**, **2l** and **2m**) were also compared for their predicted absorption, distribution, metabolism and excretion (ADME) properties with respect to other compounds previously described as G-quadruplex binders. These calculations were performed employing the QikProp

software (QikProp, version 3.4, Schrödinger, LLC, New York, NY, 2011). In addition to predicting molecular properties, QikProp provides ranges for comparing each compound's property with those of 95% of known drugs. This software was also used because it allows for flagging reactive functional groups that may cause false positives in biological assays. According to these calculations, only seven out of 32 inspected compounds display no violations of the ranges recommended for each descriptor or property. Strikingly, compounds **1a**, **2c**, **2l** and **2m** are among these most promising ligands.

7.3 Conclusions

Targeting of DNA secondary structures such as G-quadruplexes is now considered an appealing opportunity for drug intervention in anticancer therapy¹⁴. So far, efforts made in the discovery of chemotypes able to target G-quadruplexes mainly succeeded in the identification of a number of polyaromatic compounds featuring end-stacking binding properties. Unfortunately, the poor drug-like properties of these compounds turned out to be a main limitation during the *in vivo* verification of their anti-tumor properties. Herein, with the aim of discovering G-quadruplex groove binders with enhanced drug-like properties, a lead optimization campaign was undertaken starting from a promising virtual screening hit. Thus, the physico-chemical characterization (NMR and ITC) of the binding of a set of closely related analogues allowed to identify novel ligands of the [d(TGGGGT)]₄ quadruplex. Interestingly, their biological characterization demonstrated the ability to induce selective DNA damage at telomeric level and

induction of apoptosis and senescence on tumor cells. These results substantiate our choice of using the [d(TGGGGT)]₄ structure as working model to design new molecular entities endowed with G-quadruplex binding properties. Furthermore, for the first time, we demonstrate that selective G-quadruplex binding and telomeric DNA-damage can be both elicited by more drug-like chemotypes. These findings pave the way to the design of new potential drugs and shed new insights onto the emerging field of DNA quadruplex.

7.4 Materials and Methods

Oligonucleotide synthesis. The oligonucleotide d(TGGGGT) was synthesized using standard protocol¹⁵. The oligomer was detached from the support and deprotected by treatment with concentrated aqueous ammonia at 55 °C for 12 h. The combined filtrates and washings were concentrated under reduced pressure, redissolved in H₂O, analyzed and purified by high-performance liquid chromatography (HPLC) on a Nucleogel SAX column (Macherey–Nagel, 1000-8/46); using buffer A: 20 mM KH₂PO₄/K₂HPO₄ aqueous solution (pH 7.0), containing 20% (v/v) CH₃CN; buffer B: 1 M KCl, 20 mM KH₂PO₄/K₂HPO₄ aqueous solution (pH 7.0), containing 20% (v/v) CH₃CN; a linear gradient from 0 to 100% B for 30 min and flow rate 1 ml/min were used. The fractions of the oligomer were collected and successively desalted by Sep-pak cartridges (C-18). The isolated oligomer proved to be > 99% pure NMR.

Selected compounds. The selected compounds were purchased by supplier.

The purity of compounds **1a-f** and **2a-t** was assessed using reversed-phase high-performance liquid chromatography (HPLC) analyses using Shimadzu C18, 5 μ m (150 mm \times 4.6 mm) column. The elution was performed with a 1.0 ml/min flow rate using a linear gradient from 0 to 100% methanol in water over 30 min. The detection was performed at 210 nm. The purity was also tested with high-performance liquid chromatography–mass spectrometry (HPLC–MS) analyses performed on an Agilent 1200 series (Agilent Technologies, Santa Clara, CA, USA) equipped with an Agilent 6110 Series LC/MS Quadrupole, using a Phenomenex Luna[®] C18, 5 μ m (150 mm \times 4.6 mm). The elution was performed with a 1.0 ml/min flow rate using a linear gradient from 0 to 90% acetonitrile in water over 20 min. Detection was performed at 210 nm. The relative purity of compounds **1a**, **1c–e**, **2a–d**, **2f**, **2h**, **2i**, **2l–n** and **2p–r** was higher than 98.0%. Purity of compounds **1b**, **1f**, **2e**, **2g**, **2j**, **2k**, **2o**, **2s** and **2t** ranged between 95 and 98%.

Nuclear magnetic resonance experiments. The quadruplex NMR samples were prepared at a concentration of 0.1 mM (0.4 mM single strand concentration), in 0.2 ml (H₂O/D₂O 9:1) buffer solution having 10 mM KH₂PO₄, 70 mM KCl, 0.2 mM EDTA, pH 7.0. NMR spectra were recorded with Varian Unity INOVA 700 MHz spectrometer. ¹H chemical shifts were referenced relative to external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). 1D proton spectra of the sample in H₂O were recorded using pulsed-field gradient DPGSE^{16,17} for H₂O suppression. The NMR data were processed on iMAC running iNMR software (www.inmr.net).

Chemical Shift Perturbation experiments. The quadruplex [d(TGGGGT)]₄ has been titrated with each of the selected compound. The samples **1e-1f**, **2a**, **2d**, **2p-2t** were not able to cause any significant shift of the DNA signals. On the other hand, the compounds **1b-1d**, **2b**, **2c**, **2e-2k**, **2l-2o** turned out to be able to bind the quadruplex. Particularly, for all the titrations of **1b-1d**, **2b**, **2c**, **2e-2k**, **2l-2o**, the four DNA strands turned to be magnetically equivalent throughout the titration, and no splitting of resonances was observed at any stage. In order to preliminarily evaluate the binding site of each analogue, a comparison of resonances of some protons of the uncomplexed DNA and the complexed one has been done. In particular, we report the $\Delta\delta$ values (chemical shifts of the complex minus free DNA) of aromatic, methyl and imino protons. Generally, the signal of the protons of T1 residue shifted the least, whereas the ones of residue T6 the more. In any case, a general shift of the aromatic and imino signals was observed also for the G2, G3, G4 and G5.

Isothermal titration calorimetry. The d(TGGGGT) and d(CGCGAATTCGCG) oligonucleotide sequences were prepared by dissolving the lyophilized compound in 10 mM phosphate buffer with 70 mM KCl, 0.2 mM EDTA, pH 7. The solutions were annealed by heating at 90 °C for 5 min and slowly cooling to room temperature, and then equilibrated at 4 °C for 24 h. The concentration of oligonucleotides was determined by UV adsorption measurements at 90 °C using molar extinction coefficient values $\epsilon_{(260\text{ nm})}$ of 57800 and 110700 M⁻¹ cm⁻¹ for d(TGGGGT) and d(CGCGAATTCGCG), respectively. The molar extinction coefficients were calculated by the nearest neighbor model¹⁸. Stock solutions of the investigated compounds were prepared by solubilizing weighted amounts in DMSO to a final concentration

of 8 mM. The mixtures of the DNA molecules and the compounds were prepared diluting the ligand stock solution into the DNA solution to get a final ligand/DNA molar ratio of 4:1 and a final DMSO concentration of 7%. Distamycin A was solubilized in the same buffer used for the mixtures. ITC experiments were performed at 298 K using a CSC 5300 Nano-ITC microcalorimeter from Calorimetry Science Inc. (Lindon, Utah) with a cell volume of 1 ml. The titrations were carried out in 10 mM phosphate buffer, 70 mM KCl, 0.2 mM EDTA, 7% DMSO, pH 7. In each experiment, volumes of 5-10 μ L of distamycin A solution (360-720 μ M) were added into a 30 μ M solution of DNA or DNA/ligand mixture, using a computer-controlled 250 μ L microsyringe, with a spacing of 200-400 s between each injection. Each titration was corrected by subtracting the heat of distamycin A dilution. Where possible, integrated heat data obtained for the titrations were fitted employing a nonlinear least-squares minimization algorithm to a theoretical titration curve, using the Bindwork software from Calorimetry Science Inc.

Cells and culture conditions. Transformed human BJ fibroblasts expressing hTERT and SV40 early region (BJ-HELT), and human epithelial carcinoma cell line (HeLa) were obtained as previously reported⁸ and grown in Dulbecco Modified Eagle Medium (D-MEM, Invitrogen Carlsbad, CA, USA) supplemented with 10% fetal calf serum, 2 mM L-glutamin and antibiotics.

Immunofluorescence. Immunofluorescence was performed as previously reported²⁰. Cells were fixed in 2% formaldehyde and permeabilized in 0.25% Triton X100 in PBS for 5 min at room temperature. For immunolabeling experiments, cells

were incubated with primary antibody, then washed in PBS and incubated with the secondary antibodies. The following primary antibody were used: pAb anti-TRF1 (Abcam Ltd.; Cambridge UK); mAb anti- γ H2AX (Upstate, Lake Placid, NY). The following secondary antibody were used: TRITC conjugated Goat anti Rabbit, FITC conjugated Goat anti Mouse (Jackson Lab.). Fluorescence signals were recorded by using a Leica DMIRE2 microscope equipped with a Leica DFC 350FX camera and elaborated by a Leica FW4000 deconvolution software (Leica, Solms, Germany).

Clonogenic assay. HeLa cells were seeded in 60-mm Petri dishes (Nunc, MasciaBrunelli, Milan, Italy) at a density of 5×10^2 cells per dish and 24 hrs later exposed to different doses (ranging from 0.5 to 5 mM) of 2m. Cell colony-forming ability was determined as previously described²⁰. All the experiments were repeated four times in triplicate.

Flow cytometric analysis. The cell cycle analysis was performed by flow cytometry. Cells were washed in PBS and fixed in 70% ethanol in PBS. 1×10^6 cells were centrifuged and resuspended in a staining solution (50 μ g/ml PI, 75 KU/ml RNase A in PBS) for 30 min at room temperature in the dark and analyzed by flow cytometry using FACScalibur (Becton-Dickinson, San Jose, CA, USA). For each analysis 20.000 events were collected. Cell cycle distribution and percentage of apoptotic cells was analyzed using Cell Quest (BDIS) and ModFit LT (Verity Software House, Topsham, ME).

Evaluation of apoptosis. Apoptosis was detected by flow cytometric analysis of annexin V staining. Annexin V–FITC vs PI assay (Vibrant apoptosis assay, V-13242, Molecular Probes, Eugene, OR, USA) was performed as previously described²¹. Briefly, adherent cells were harvested and suspended in the annexin-binding buffer (1×10^6 cells/ml). Thereafter, cells were incubated with annexin V–FITC and PI for 15 min at room temperature in the dark and immediately analyzed by flow-cytometry. The data are presented as bi-parametric dot plots showing PI red fluorescence vs annexin V–FITC green fluorescence.

Senescence Analysis. Senescence-associated β -galactosidase (SA- β -gal) staining on HeLa cells was performed as described by Dimri *et al.*²² Briefly, after exposure with 3.5 μ M 2 m for 5 days of cell culture, the cells were fixed with 2% glutaraldehyde in PBS for 5 min at room temperature, washed in PBS, and incubated at 37 °C for 24 hrs in staining solution: 1 mg/ml 5-bromo-4-chloro-3-indolyl- β -d-galactoside (X-gal), 5 mm potassium ferrocyanide, 5 mm potassium ferricyanide, 2 mm $MgCl_2$ in PBS, pH 6.0. Then cells were analyzed using an optical microscope.

Statistical analysis. The experiments have been repeated from three to five times and the results obtained are presented as means \pm SD. Significant changes were assessed by using Student's t test for unpaired data, and P values < 0.05 were considered significant.

References

- [1] Cosconati S., Marinelli L., Trotta R., Virno A., Mayol L., Novellino E., Olson A. J., Randazzo A. *J. Am. Chem. Soc.* **2009**, 131, 16336-16337.
- [2] a) Li Q., Xiang J., Li X., Chen L., Xu X., Tang Y., Zhou Q., Li L., Zhang H., Sun H., Guan A., Yang Q., Yang S., Xu G. *Biochimie* **2009**, 91, 811-829; b) Ranjan N., Andreasen K. F. Kumar S., Hyde-Volpe D., Arya D. P. *Biochemistry* **2010**, 49, 9891-9903; c) Hamon F., Largy E., Guédin-Beaurepaire A., Rouchon-Dagois M., Sidibe A., Monchaud D., Mergny J. L., Riou J. F., Nguyen C. H., Teulade-Fichou M. P. *Angew. Chem. Int. Ed.* **2011**, 50, 8745-8749.
- [3] Trotta R., De Tito S., Lauri I., La Pietra V., Marinelli L., Cosconati S., Martino L., Conte M. R., Mayol L., Novellino E., Randazzo A. *Biochimie* **2011**, 93, 1280-1287.
- [4] a) Randazzo A., Galeone A., Mayol L. *Chem. Comm.* **2001**, 11, 1030-1031; b) Randazzo A., Galeone A., Esposito V., Varra M., Mayol L. *Nucleoside Nucleotides Nucleic Acids* **2002**, 21, 535-545; c) Martino L., Virno A., Pagano B., Virgilio A., Di Micco S., Galeone A., Giancola C., Bifulco G., Mayol L., Randazzo A. *J. Am. Chem. Soc.* **2007**, 129, 16048-16056; d) Pagano B., Virno A., Mattia C. A., Mayol L., Randazzo A., Giancola C. *Biochimie* **2008**, 90, 1224-12232; e) Cosconati S., Marinelli L., Trotta R., Virno A., De Tito S., Romagnoli R., Pagano B., Limongelli V., Giancola C., Baraldi P.G., Mayol L., Novellino E., Randazzo A. *J. Am. Chem. Soc.* **2010**, 132, 6425-6433.
- [5] Borges F., Roleira F., Milhazes N., Santana L., Uriarte E. *Curr. Med. Chem.* **2005**, 12, 887-916.
- [6] Petraccone L., Fotticchia I., Cummaro A., Pagano B., Ginnari-Satriani L., Haider S., Randazzo A., Novellino E., Neidle S., Giancola C. *Biochimie* **2011**, 93, 1318-1327.
- [7] Pagano B., Cosconati S., Gabelica V., Petraccone L., De Tito S., Marinelli L., La Pietra V., Di Leva F. S., Lauri I., Trotta R., Novellino E., Giancola C., Randazzo A. *Curr. Pharm. Des.* **2012**, 18, 1880-1899.
- [8] Horton D. A., Bourne G. T., Smythe M. L. *Chem. Rev.* **2003**, 103, 893-930.
- [9] Velazquez Campoy A., Freire E. *Biophys. Chem.* **2005**, 115, 115-124.

- [10] a) Pagano B., Mattia C. A., Virno A., Randazzo A., Mayol L., Giancola C. *Nucleosides Nucleotides Nucleic Acids*, **2007**, 26, 761-765. b) Pagano, B.; Mattia, C. A., Giancola C. *Int. J. Mol. Sci.* **2009**, 10, 2935-2957.
- [11] Klevit R.,E., Wemmer D.,E., Reid B.,R. *Biochemistry* **1986**, 25, 3296-3303.
- [12] van Steensel B., de Lange T. *Nature* **1997**, 385, 740-743.
- [13] Takai H., Smogorzewska A., de Lange T. *Curr. Biol.* **2003**, 13, 1549-1556.
- [14] a) Yu H., Wang X., Fu M., Ren J., Qu X. *Nucleic Acids Res.* **2008**, 36, 5695-5703; b) Yu H., Zhao C., Chen Y., Fu M., Ren J., Qu X. *J. Med. Chem.* **2010**, 53, 492-498.
- [15] For example, see: a) D'Onofrio J., Petraccone L., Martino L., Di Fabio G., Iadonisi A., Balzarini J., Giancola C., Montesarchio D. *Bioconjug. Chem.* **2008**, 19, 607–616; b) Di Fabio G., D'Onofrio J., Chiapparelli M., Hoorelbeke B., Montesarchio D., Balzarini J., De Napoli L. *Chem. Commun.* **2011**, 47, 2363–2365.
- [16] Hwang T. L., Shaka A. J. *J. Magn. Reson.* **1995**, A112, 275–279.
- [17] Dalvit C. *J. Biomol. NMR* **1998**, 11, 437–444.
- [18] Cantor C. R., Warshaw M.M., Shapiro H. *Biopolymers* **1970**, 9, 1059-1077.
- [19] a) Gomez D., O'Donohue M.F., Wenner T., Douarre C., Macadré J., Koebel P., Giraud-Panis M. J., Kaplan H., Kolkes A., Shin-ya K., Riou J. F. *Cancer Res.* **2006**, 66, 6908-6912; b) Pagano B., Giancola C. *Curr. Cancer Drug Targets* **2007**, 7, 520-540; c) Salvati E., Leonetti C., Rizzo A., Scarsella M., Mottolese M., Galati R., Sperduti I., Stevens M. F., D'Incalci M., Blasco M., Chiorino G., Bauwens S., Horard B., Gilson E., Stoppacciaro A., Zupi G., Biroccio A. *J. Clin. Invest.* **2007**, 117, 3236-3247.
- [20] Biroccio A., Amodei S., Benassi B., Scarsella M., Cianciulli A., Mottolese M., Del Bufalo D., Leonetti C., Zupi G. *Oncogene* **2002**, 21, 3011-3019.
- [21] Biroccio A., Benassi B., Filomeni G., Amodei S., Marchini S., Chiorino G., Rotilio G., Zupi G., Ciriolo M. R. *J. Biol. Chem.* **2002**, 277, 43763-43770.
- [22] Dimri G. P., Lee X., Basile G., Acosta M., Scott G., Roskelley C., Medrano E. E., Linskens M., Rubelj I., Pereira-Smith O., Peacocke M., Campisi J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, 92, 9363–9367.

CHAPTER IV

**INSIGHT INTO NEW DNA TOPOLOGIES: THE
G-TRIPLEX**

1. The G-Triplex DNA

1.1 Introduction

Nucleic acids represent the alphabet of the cellular language generating through their sequencing encrypted messages that regulate functions vital for the cell. In 1953, the discovery that DNA could adopt a double helix conformation¹ (Figure 1.1), represented a milestone in biology, showing that important genetic information could be delivered by specific DNA topology.

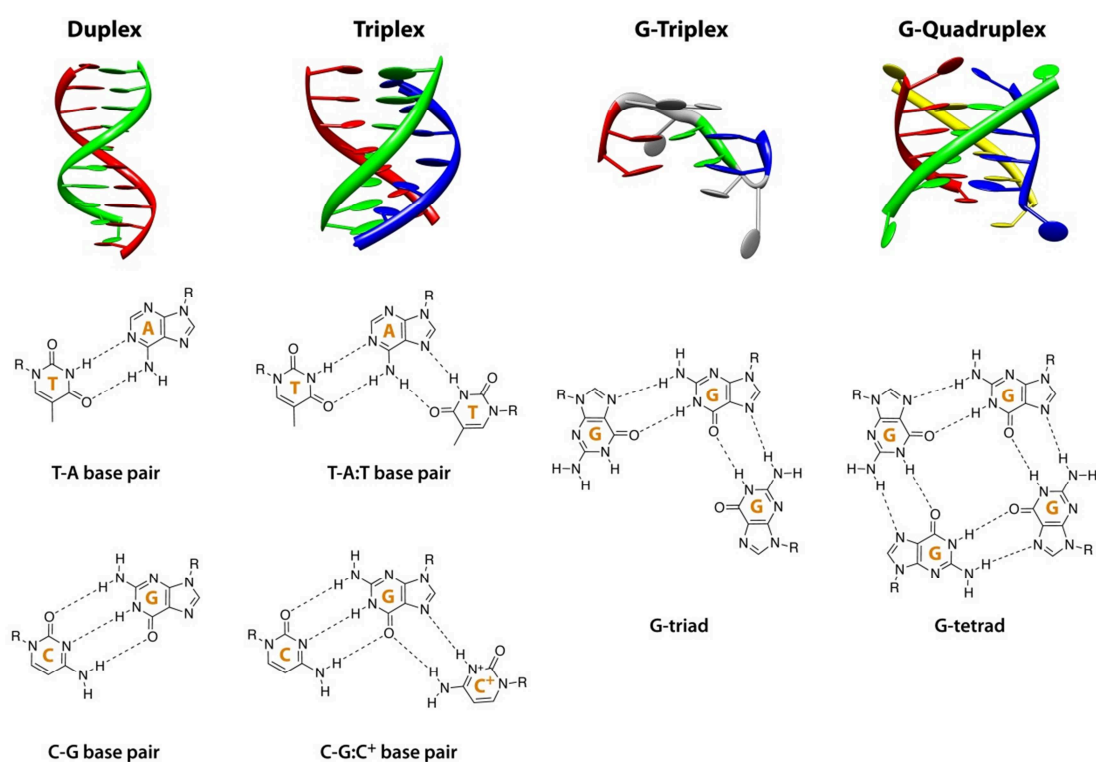


Figure 1.1: DNA structural motifs. (Top panel) Schematic illustration of duplex, triplex, G-triplex and G-quadruplex. (Bottom panel) Examples of base pairing: T-A for duplex, T-A:A and C-G:C⁺ for triplex (where G-C and A-T indicates the regular Watson-Crick pairing, whereas the colon signs indicate the pairing with the third strand), G-triad for G-triplex, and G-tetrad for G-Quadruplex.

The specificity and the complexity of the genetic language were further increased in recent years, since variations from the Watson-Crick duplex structure have been found to play key roles in many cellular processes. Examples are higher-order architectures of DNA, such as hairpins², cruciforms³, parallel-stranded duplexes⁴, triplexes⁵, G-quadruplexes⁶ and the i-motif⁷, that are involved in several important processes like DNA packaging, replication, transcription and recombination. These structures can be formed by nucleotides sequences distributed throughout the whole human genome and their location in the genome is not random but associated with human diseases⁸. From a structural point of view, all these complexes are formed by one up to four DNA strands, stabilized by base stacking and base pairing interactions with a wider variety of base pairing motifs than the standard Watson-Crick ones (Figure 1.1). For instance, DNA triplexes can present G:G-C, A:A-T, C⁺:G-C and T:A-T base pairings with two strands in the standard Watson-Crick duplex structure (*i.e.* G-C and A-T) and the third one lying in the major groove of the duplex (Figure 1.1). Another example are G-quadruplexes, that are four stranded structures and have a core of two or more guanine tetrads stacked one upon the other. A G-tetrad consists of a planar arrangement of four guanine bases associated through a cyclic array of Hoogsteen-like hydrogen-bonds in which each guanine accepts and donates two hydrogen-bonds (Figure 1.1). Nucleotide sequences able to adopt structures like triplexes and G-quadruplexes, are very abundant in the human genome and play key roles in the regulation of many genes functions vital for the cell⁹.

These findings have highlighted the high polymorphism of the genetic language suggesting that there might exist other DNA structures with specific

cellular functions that are to date unknown. In this scenario, the discovery of novel DNA structural motifs is more than ever an important breakthrough that provides new molecular bases to elucidate functional mechanisms of cell otherwise unexplained and unveil the way to interact with them.

Here, combining advanced state-of-the-art computations and experiments, we have identified a new DNA structural motif. This structure is named “G-triplex” and at variance with the known triplexes, it is formed by G:G:G triad planes stabilized by an array of Hoogsteen-like hydrogen-bonds (Figure 1.1). This is the first time that DNA is found to assume this kind of topology and unveiling its biological role, which is to date unknown, represents a major research goal in the near future.

1.2 Results and Discussion

Metadynamics study. The biological relevance of DNA structures like triplexes and G-quadruplexes, makes imperative to elucidate their functional mechanism and new insight in this sense can be obtained by the study at molecular level of their folding mechanism. Unfortunately, this kind of processes takes from microseconds to hundreds of seconds, thus is impossible to simulate in the time-scale of standard computational techniques such as molecular dynamics (MD). However, advances in the study of large time-scale processes, have been recently reported with the aid of enhanced sampling techniques such as metadynamics¹⁰⁻¹². In fact, metadynamics has been successfully used even in complex ligand docking studies, such as the cyclooxygenases case¹³, being able to predict the binding conformation of a potent inhibitor, SC-558, in COX-1, prediction that was

subsequently validated by X-ray experiments¹⁴. This technique enhances sampling by adding a bias on few degrees of freedom of the system, called collective variables (CVs), allowing to investigate large time-scale processes in affordable computational time¹⁵. Furthermore, metadynamics allows to reconstruct the free energy profile, thus identifying different conformations of the system and the energetic cost of converting one into the other. This is of fundamental importance in the present case study, since ordered DNA structures may form transiently at specific times during the cell cycle with a specific role and regulatory mechanism.

While many studies on protein folding are reported in literature, DNA/RNA folding studies are still in their infancy thus representing a cutting edge topic of research. We decided to tackle this challenge and using well-tempered metadynamics simulations¹⁶, we have investigated the folding/unfolding process of one of the simplest G-quadruplex structure: the Thrombin Binding Aptamer (TBA)¹⁷. TBA is formed by 15 nucleotides (5'-dGTTGGTGTGGTTGG-3') organized in an anti-parallel monomolecular G-quadruplex, characterized by a chair-like structure consisting of two G-tetrads connected by two TT loops and a single TGT loop (Figure 1.1). We have used as CV the radius of gyration formed by the oxygen atoms of the guanine bases of the G-quadruplex planes (see Methods for details). This CV is able to describe the formation and the disruption of the G-tetrad planes and has been already used in other metadynamics studies of protein folding^{18,19}. Although this CV can well describe the global motion of the system, it is not able to distinguish the multiple conformations assumed by the system during the folding process. In order to overcome this limitation and have a clearer representation of the free energy profile associated with the TBA folding process, we have computed the Free Energy

Surface (FES) as a function of a second CV, namely the number of hydrogen-bonds formed by the guanine bases of the G-tetrads (see Methods for details). This can be done using the reweighting algorithm²⁰ recently developed in our lab (see Methods for details). Looking at the FES obtained after around 80 ns of metadynamics simulation, three main energy minima can be identified (Figure 1.2).

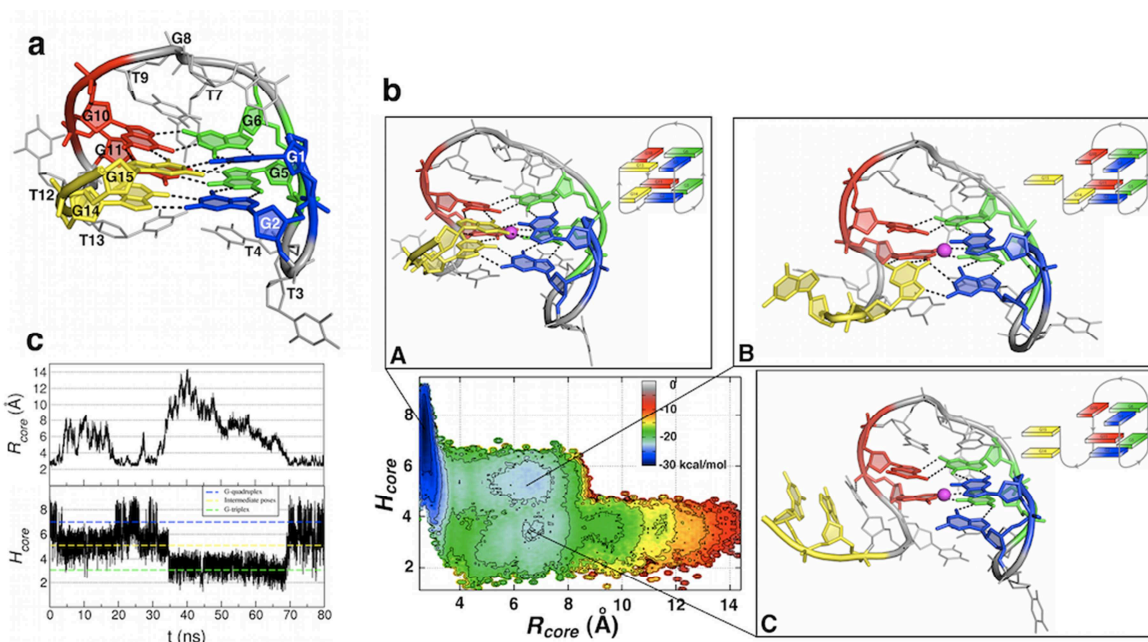


Figure 1.2: Representation of the Free Energy Surface (FES) of the 3' end opening of TBA. a, Representation of the NMR structure of TBA (PDB ID code 1qdf). b, In the FES three main energy minima are shown: one deep and narrow, basin A, which represents TBA in the G-quadruplex structure; the second one, basin B, which represents an intermediate state; and the last one, basin C, which shows the G-triplex structure formed by the triads G1:G6:G10 and G2:G5:G11. c, Plots showing the phase space represented as the H_{core} and R_{core} CV, explored during the metadynamics simulation. It can be observed that the 3' end of TBA opens and closes several times, passing from basin A to basin B and basin C, and then folded again in the G-quadruplex structure, basin A. Thanks to the recrossing events between different basins, the calculated FES is accurate and quantitatively well characterized.

The deepest one, basin A, is at the gyration radius CV (CV1) value of 2.5 Å and at high values of the Hydrogen-bonds CV (CV2). This pose corresponds to the experimental G-quadruplex structure of TBA²¹ with two G-tetrad planes and eight

guanines involved in standard Hoogsteen hydrogen-bonds. Interestingly, during the metadynamics simulation the three loops have a different behavior. While the GTG loop is rather stable, the two TT loops show higher flexibility with T3 and T12 changing their orientation several times during the simulation. At variance with T3 and T12, T4 and T13 conserve their original position showing a good conformational stability. This is in line with the NMR data, to be described below, that indicate that these bases are involved in hydrogen-bond interactions. In our case, T4 and T13, albeit very close, do not form hydrogen-bond interactions, while stacking interactions with the guanines placed above, G5 and G14 for T4 and T13, respectively, are engaged and contribute to the energy stability of this conformation. Looking at the FES two further minima, basin B and basin C, can be detected at higher energy values. In basin B, TBA shows a partial opening of the 3' end with G15 oriented towards the solvent. Interestingly, in this basin G14 moves slightly from its original position pointing the oxygen of the nucleobase towards the center of the planes formed by G1:G6:G10 and G2:G5:G11 (Figure 1.2). In such a way, G14 conserves the hydrogen-bond interactions with G2 and G11 and at the same time it partially fills the vacancy in the coordination shell of the metal present at the core of the structure. This step can be considered the very first event in the unfolding process of TBA.

G-triplex structure. After the partial opening of the 3' end, TBA folded again into its original G-quadruplex structure (Figure 1.2). Subsequently, the 3' end opens completely with G14 and G15 leaving the G-tetrad planes and pointing towards the solvent. This state is represented by the broad energy minimum C in Figure 1.2. This

minimum is approximately 6.5 kcal/mol higher in energy than basin A and here TBA assumes a number of different conformations due to the conformational flexibility of the 3' overhang formed by T13-G14-G15. These bases rearrange to form a single helix that assumes a number of conformations, all of these stabilised by stacking interactions between the bases. While the 3' overhang is flexible, the rest of the TBA molecule is rather stable forming two G:G:G planes, namely G-triads, composed by G1:G6:G10 and G2:G5:G11, that engage an array of Hoogsteen-like hydrogen-bonds (Figure 1.1). In this conformation the metal is coordinated at the center of the two triads in a way similar to that of the G-quadruplex structure. This is the first time that DNA is found in a triplex conformation with G:G:G base pairs. This structure is named "G-triplex" to distinguish it from the known triplex structures that present G:G-C, A:A-T, C⁺:G-C and T:A-T base pairings with two strands in the standard Watson-Crick duplex structure and the third one lying in the major groove of the duplex (Figure 1.1). At variance with the other two minima A and B, in basin C, T12 does not point out towards the solvent, on the contrary it is rotated towards the core of TBA engaging stacking interactions with T4 of the TT loop at the 5' end (basin C in Figure 1.2). This position favors the formation of hydrogen-bond interactions between the imide hydrogens of T13 and G14 and the backbone phosphates of G11. These interactions further contribute to the energy stability of this pose. The stability of the TBA conformation in the three energy minima A, B and C, has been further assessed through standard molecular dynamics calculations.

Molecular dynamics simulations on TBA in the conformations representing the energy basins A, B and C, and on the oligonucleotide 4. We assessed the

stability of the conformations representing the energy basins A, B and C using 50 ns long molecular dynamics simulations. Different behaviors have been observed for the different poses considered. As expected, basin A resulted very stable throughout the whole simulation with the two G-tetrads planes formed by G1:G6:G10:G15 and G2:G5:G11:G14, constantly formed (Figure 1.3).

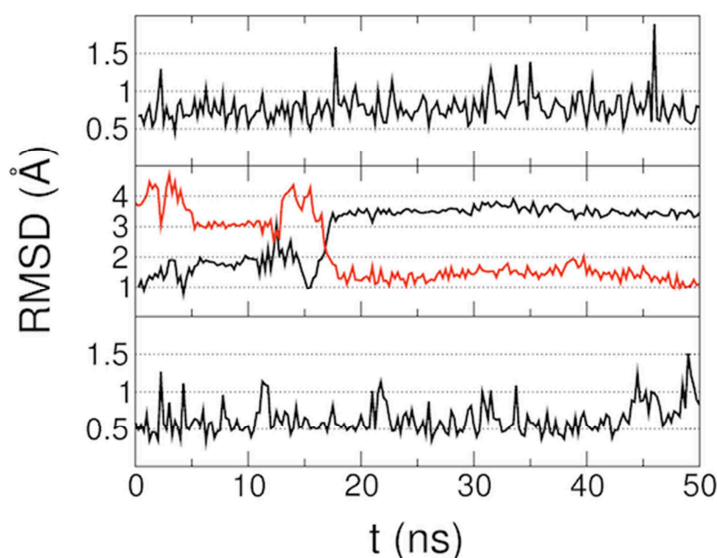


Figure 1.3: Plots regarding the MD simulations on basin A, B and C conformation. (Top graph) Plot of the rmsd of the heavy atoms of guanines forming the G-tetrad planes, G1:G6:G10:G15 and G2:G5:G11:G14, during over 50 ns of MD simulation. The very low rmsd values reflect the good stability of this pose. (Middle graph) Plot of the rmsd of the heavy atoms of guanines forming the G-tetrad planes, G1:G6:G10:G15 and G2:G5:G11:G14, during over 50 ns of MD simulation using the basin B pose as starting conformation. The rmsd are calculated relatively to the basin B (black lines) and basin A (red lines) conformation. The red plots clearly show that the basin B conformation changes in the basin A one after approximately 17 ns. This change is possible thanks to the relatively low energy barrier that separates basin B from basin A. (Lower graph) Plot of the rmsd of the heavy atoms of guanines forming the G-triad planes, G1:G6:G10 and G2:G5:G11, during over 50 ns of MD. The very low average rmsd value of 0.63 Å reflects the high stability of the G-triplex structure.

As seen in the metadynamics calculation, even in the MD simulation the T7-G8-T9 loop is rather stable while T12 in the T12-T13 loop, is very flexible and points away towards the solvent. Furthermore, T3, T4 and T13 present a good conformational stability favored by the engagement of stacking interactions with the above placed nucleobases.

At variance with basin A, basin B is rather unstable. At this basin the partial opening of the 3' end of TBA has been observed during the metadynamics simulation, with the G15 base pointing towards the solvent (Figure 1.2). As can be seen in Figure 1.3, after approximately 18 ns of MD simulation TBA leaves its starting conformation, basin B, transforming in that of basin A. In particular, G15 slowly moves towards the center of the G-tetrads forming again the G-tetrads planes and the Hoogsteen hydrogen-bonds first with G1 and then with G10. This finding is not totally surprising since, looking at the FES of the metadynamics simulation (Figure 1.2), the energy barrier that separates basin B from basin A, is rather small, approximately 2 kcal/mol, and thus it can be overcome in the standard MD time-scale. This motion can be better appreciated looking at Figure 1.3 where the r.m.s.d. of the guanine bases with respect to their position in the G-quadruplex conformation has been calculated.

A different behavior has been instead observed for basin C. In fact, during the MD simulation the energy minimum TBA conformation is stable conserving all the interactions previously found in the metadynamics calculation. In particular, at this pose, the 3' end of TBA is completely open with T13, G14 and G15 forming a single helix overhang that points away from the core of the TBA structure (Figure 1.2). The other nucleobases are arranged to form a well-organized structure where G1:G6:G11 and G2:G5:G10 form two G-triad planes stabilized by an array of Hoogsteen-like hydrogen-bonds. This structure, namely G-triplex, is very stable along the whole MD simulation with a low average r.m.s.d. of 0.63 Å for the guanine bases forming the G-triad planes (Figure 1.3).

Furthermore, we have carried out a very long MD simulation, more than 100 ns, on the oligonucleotide **4** in its G-triplex conformation as found by the NMR experiments. This calculation has been performed simulating the experimental buffer condition of 70 mM KCl. During this simulation the G-triplex structure is very stable with a low average r.m.s.d. of 0.88 Å for the guanines forming the triad planes. In fact, the triads formed by G1:G6:G11 and G2:G5:G10 are stabilized by Hoogsteen-like hydrogen-bonds as previously described in the other simulations (Figure 1.4).

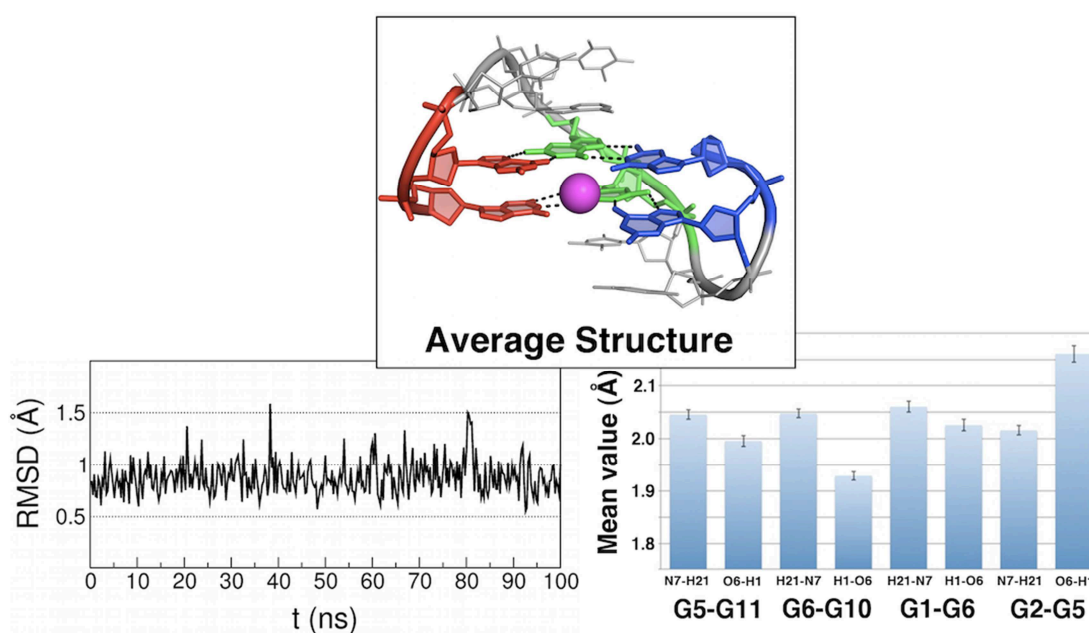


Figure 1.4: Plots regarding the MD simulation of the oligonucleotide **4** represented as average structure. (Left graph) Plot of the rmsd of the heavy atoms of guanines forming the G-triad planes, G1:G6:G10 and G2:G5:G11, in the 11-bases oligonucleotide (**4**). The very low average rmsd value of 0.88 Å reflects the high stability of the G-triplex structure that is represented in the inset picture. (Right graph) Representation as histograms of the mean values of the distances between the hydrogen-bond acceptor and hydrogen-bond donor groups in guanines forming G-triplex. The error bars show the SEM (standard error of the mean). Data for both the graphs are obtained from the over 100 ns MD simulation reproducing the experimental ion concentration.

At the center of the two triad planes the potassium ion is coordinated by six oxygen atoms of the guanines. Analogously to the metadynamics results, T4 has a

good conformational stability thanks to the stacking interactions engaged with G5, while T3 is more flexible. However, during the simulation this base often occupies a position under G2 where favorable stacking interactions can be formed. On the other hand, the T7-G8-T9 loop is stable with G8 stacked above the plane formed by the G1:G6:G10 triad engaging favorable interactions with G6.

All the simulations have shown a great stability of the G-triplex structure in both TBA and oligonucleotide **4**, suggesting to investigate the presence of G-triplex structures even in other poliguanine sequences of DNA and RNA.

Nuclear Magnetic Resonance. In order to provide experimental evidence of the formation of the G-triplex structure, four oligodeoxynucleotides formed by the gradual truncation of the 3' end of the TBA sequence (5'-GGTTGGTGTGGTTGG-3') have been investigated in solution by NMR: 5'-GGTTGGTGTGGTTG-3' (**1**), 5'-GGTTGGTGTGGTT-3' (**2**), 5'-GGTTGGTGTGGT-3' (**3**) and 5'-GGTTGGTGTGG-3' (**4**). Samples **1-4** were analyzed by ^1H -NMR at 25 °C in 80 mM K^+ -containing buffer (see experimental section). The spectra of samples **1**, **2** and **3** exhibited conformational heterogeneity (data not shown). This does not depend on the nature of the buffer solution, be it a potassium or sodium containing buffer, or on its concentration and temperature. At variance with these previous results, the analysis of the oligodeoxynucleotide 5'-GGTTGGTGTGG-3' (**4**) clearly indicates the presence of a predominant well-defined hydrogen-bonded structure in solution, proven by the presence of four well defined exchangeable proton signals in the 11.0 - 12.5 ppm region of the 1D ^1H -NMR spectrum (Figure 1.5a).

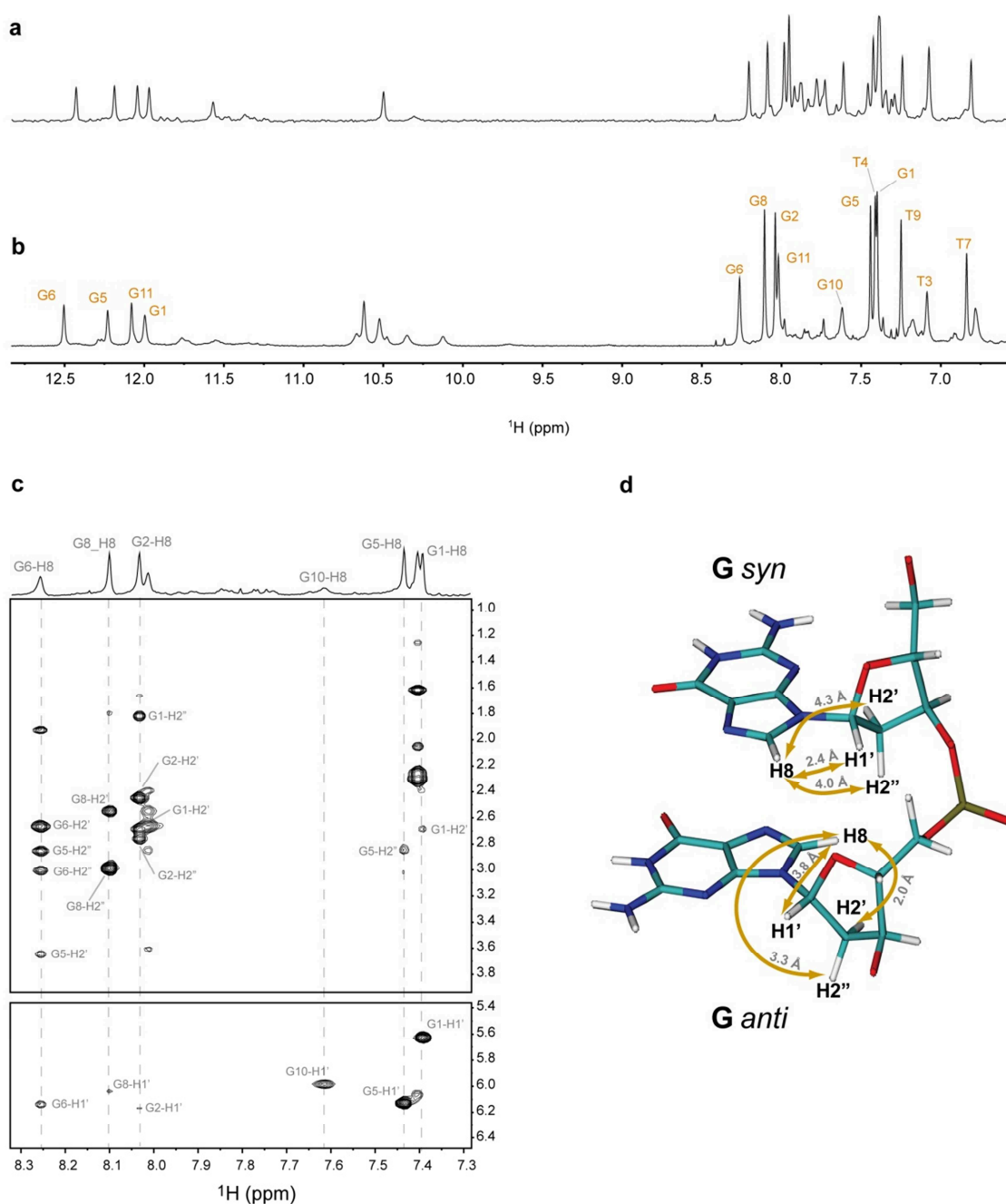


Figure 1.5: 1D ^1H -NMR and 2D NOESY spectra of 5'-GGTTGGTGTGG-3' (**4**). Imino, amino and aromatic regions of the 1D ^1H -NMR spectra of **4** acquired at 25 °C (**a**) and 1 °C (**b**). (**c**) Two expanded regions of the NOESY spectrum (900 MHz, T = 1 °C, mixing time 100 ms). H8 protons of the *syn* residues G1, G5 and G10 show intense cross-peaks with H1' proton and weak connectivities with H2' and H2'' protons of their own sugar. On the contrary, *anti* residues G2, G6, G8 have opposite relative intensity. (**d**) Distances of the correlated protons in the NOESY spectrum.

It is worth noting that signals that resonate in this region, are typical diagnostic of the formation of DNA structures presenting Hoogsteen Hydrogen-bonds network²². On the other hand, the region of aromatic protons (6,5 - 8,5 ppm)

of **4** is characterized by the presence of eleven intense signals that can be attributed to the seven guanine H8 and four thymine H6 protons, and by a number of minor signals (Figure 1.5a). The presence of these minor signals in the aromatic region could be attributed to the unstructured DNA that is in equilibrium with the structured one. The equilibrium between folded (structured) and unfolded (unstructured) forms turns out to be very sensitive to temperature. In fact, the intensity of the signals of the two species can be changed by varying the temperature, as clearly demonstrated by the spectrum acquired at 1 °C, in which the intensities of the minor signals are <10% of the major species (Figure 1.5b). Interestingly, in the latter spectrum the signal at 7.63 ppm, attributed to G10-H8 (see later), is broad, suggesting that its conformation varies on the NMR timescale.

Thermodynamic characterization. Circular Dichroism (CD) spectroscopy has been exploited to further investigate the structure adopted by **4** in solution. CD spectra were recorded at 1 °C, a temperature at which the folded-unfolded equilibrium of **4** is strongly shifted towards the folded form, as shown by NMR, CD melting and DSC experiments. For DNA, a CD spectrum is generated by a chiral disposal of the chromophores, *i.e.* the nucleobases, and it is strictly related to the base-stacking pattern. Since the two faces of a base are heterotopic, when two bases are stacked together, they can interact through the same (head-to-head or tail-to-tail) or the opposite (head-to-tail) faces, leading to a heteropolar or homopolar stacking, respectively. When the glycosyl bonds of the guanines alternate in *syn* and *anti* conformations along the strand, head-to-head and/or tail-to-tail interactions are realized, leading to a CD signal characterized by two positive

bands at approx. 295 and 245 nm and two negative bands at approx. 230 and 265 nm^{17,18}. The experimental CD spectrum of **4** is characterized by two positive CD bands at 289 and 253 nm, and two negative bands at 235 and 265 nm (Figure 1.6a), values that are consistent with the structure proposed by metadynamics calculations and NMR^{23,24}. Such spectrum closely resembles that of TBA. However, both positive and negative bands are slightly shifted, thus suggesting a similar, but obviously not identical, stacking of the bases. Rationalizing the differences is difficult at this stage, although they may be ascribable to the fact that in **4** G1-G2 and G10-G11 strands are closer than in TBA, resulting therefore in a different stacking of the bases.

In order to investigate the thermodynamic stability and to further characterize the molecularity of the structure formed by **4**, we acquired CD spectra and performed CD melting and Differential Scanning Calorimetry (DSC) experiments.

Figure 1.6b shows the CD melting profile of **4** recorded at the wavelength of maximum absorbance variation upon folding ($\lambda = 289$ nm). As shown by almost superimposable heating and cooling curves, the melting profile was reversible, with no significant hysteresis observed, indicating a *quasi*-equilibrium process. Analysis of CD melting curves yielded a melting temperature, T_m , of 33.5 (± 1.0) °C. The model-dependent van't Hoff analysis of melting curves gave an enthalpy change, $\Delta H^\circ_{\text{vH}}$, of 145 (± 15) kJ mol⁻¹.

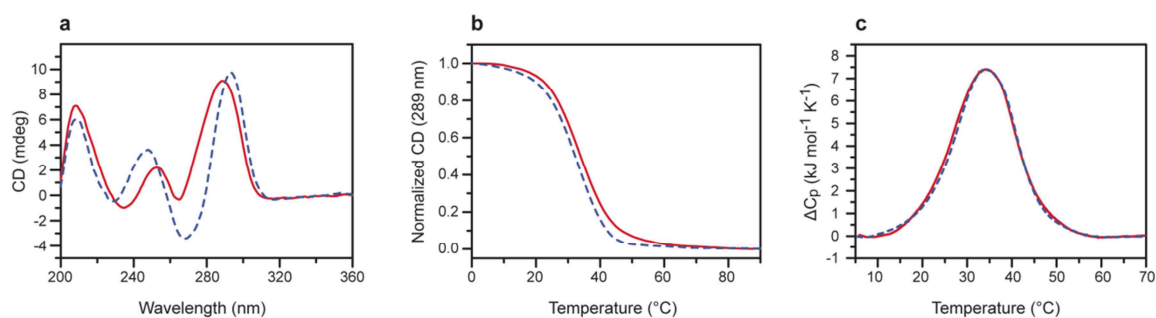


Figure 1.6: Biophysical characterization of 5'-GGTTGGTGTGG-3' (**4**). (a) CD spectra of **4** (solid line) and TBA (dashed line) at 1 °C. (b) Normalized CD melting curves of **4** recorded at 289 nm at a scan rate of 0.5 °C min⁻¹. Heating and cooling curves are shown with solid and dashed line, respectively. (c) DSC profiles for **4** at 0.5 (solid line) and 1.0 (dashed line) °C min⁻¹ heating rate. All the experiments were performed in a buffer solution containing 10 mM potassium phosphate, 70 mM KCl, 0.2 mM EDTA (pH 7.0).

DSC measurements were then carried out under the same solution conditions as the CD experiments, in an effort to characterize the denaturation thermodynamics of **4** with a model-independent analysis method²⁵.

Typical DSC thermograms for denaturation of **4** are shown in Figure 1.6c. The unfolding of **4** is a highly reversible process, as demonstrated by the recovery of the original signal by rescanning the same sample.

Furthermore, the change of the heating rate from 0.5 to 1.0 °C min⁻¹ does not alter the thermodynamic parameters significantly, thereby demonstrating that the investigated process is not kinetically controlled²⁶. The DSC curves show a symmetric shape with a maximum centered at T_m of 34.0 (± 0.5) °C, in good agreement with that obtained by CD melting. The melting temperature was almost concentration-independent, consistent with a structure resulting from unimolecular folding. The initial and final states of the investigated transition have similar heat capacity values and indicate that the unfolding of **4** is accompanied by a negligible heat capacity change. The integration of the denaturation peak gives a ΔH°_{cal} of 135 (± 5) kJ mol⁻¹. Interestingly, the change in enthalpy as obtained from the area under

the DSC curve is almost identical to the van't Hoff enthalpy calculated from DSC curves and very close to the one calculated from CD curves. This indicates that the transition of **4** takes place in a two-state manner, *i.e.*, no intermediate states are present. Finally, the Gibbs energy value, which rigorously measures the thermodynamic stability, was calculated by the experimental thermodynamic parameters. The calculated $\Delta G^\circ(298\text{ K})$ value is $4 (\pm 1) \text{ kJ mol}^{-1}$, and results from the characteristic compensation of the favorable enthalpy term with an unfavorable entropy contribution ($\Delta S^\circ = 0.44 \pm 0.02 \text{ kJ mol}^{-1}\text{K}^{-1}$). The whole set of thermodynamic parameters shows that, as expected, **4** is less stable than TBA²⁷. Particularly, the favorable enthalpy term is lower than the one derived for TBA, probably due to the greater contribution from stronger base stacking and more Hoogsteen hydrogen-bonds involved in the TBA quadruplex structure. This is consistent with the lack of G-tetrads in the structure adopted by **4**.

Structural characterization. The non-exchangeable base and sugar protons of **4** were assigned through a combination of the analysis of 2D NOESY (900 MHz, $T = 1\text{ }^\circ\text{C}$), 2D TOCSY (700 MHz, $T = 1\text{ }^\circ\text{C}$) and 2D COSY (700 MHz, $T = 1\text{ }^\circ\text{C}$) NMR spectra (see Table 1.1).

Interestingly, the presence of three intense cross-peaks between the H8 proton bases and sugar H1' resonances for residues G1, G5 and G10 in the NOESY spectrum (900 MHz, $T = 1\text{ }^\circ\text{C}$, mixing time 100 ms), along with the presence of weak cross-peaks between the same aromatic protons and the H2' and H2'' protons of their own ribose moiety, indicated that these three residues adopt a *syn* glycosidic angle conformation (Figure 1.5c). On the contrary, residues G2, G6, G8 and G11

turned out to assume an *anti* glycosidic conformation, having opposite relative intensity with respect to the cross-peaks mentioned above (Figure 1.5c). The three H8 resonances of *syn* G residues are downfield shifted with respect to those of the *anti* ones, exactly as reported for TBA²⁸.

	H8/H6	CH ₃	H1'	H2'	H2''	H3'	H4'	H5'/H5''	H1/H3	H2 (H-bonded)	H2 (non H-bonded)
G1	7.40	--	5.62	2.68	1.81	4.67	3.60	3.40, 2.86	12.00	9.72	6.15
G2	8.04	--	6.16	2.44	2.75	4.92	4.41	4.80, 4.02	12.42	6.90	6.90
T3	7.09	1.65	5.04	1.30	2.04	4.41	4.29	4.23, 3.83	--	--	--
T4	7.41	1.61	6.05	2.23	2.29	4.45	3.01	2.81, 2.59	10.34	--	--
G5	7.44	6.12	6.12	3.64	2.84	4.87	4.51	4.37, 3.96	12.23	7.17	7.17
G6	8.26	--	6.13	2.66	3.00	5.02	4.57	4.25	12.51	6.91	6.91
T7	6.86	1.92	5.50	0.50	1.79	4.65	4.17	4.00, 3.99	11.76	--	--
G8	8.11	--	6.03	2.54	2.98	5.01	4.18	4.06, 4.00	10.65	--	5.90
T9	7.25	1.54	5.91	1.72	2.22	4.52	3.62	3.44, 2.54	12.42	--	--
G10	7.63	--	5.98	2.84	2.53	4.72	4.31	3.95	11.55	6.48	6.48
G11	8.02	--	5.74	2.38	2.65	4.78	4.15	4.10, 3.60	12.08	10.12	6.91

Table 1.1: ¹H-NMR assignment of **4** at 1 °C (900 MHz)

Furthermore, the NOESY spectrum showed that the three *anti*-Gs (G2, G6 and G11) have classical H8/H2'–H2'' sequential connectivities to 5' neighboring *syn*-Gs (G1, G5, G10, respectively). This suggests the presence of the subunits G1–G2, G5–G6, G10–G11 (where underlined residues adopt a *syn* glycosidic conformation), and indicated that these subunits are involved in the formation of a helical structure. Notably, unusual NOE connectivities were observed between a number of Gs and Ts, indicating that 5'-TG-3' and 5'-GT-3' tracts do not adopt a helical winding, and that the TT and TGT tracts form loops. Furthermore, the alternation of *syn* and *anti* G residues implies that our sequence, as in TBA, folds back into a

unimolecular structure, as also demonstrated by CD and DSC experiments (see thermodynamic characterization). Since the number of imino protons is four, presumably sample **4** adopts the folding proposed by metadynamics calculations, where the presence of two G-triads is consistent with four imino protons in slow exchange with the solvent.

The assignment of the exchangeable imino protons was instead obtained by the study of the JR-HMBC correlation experiment. The jump-return version of HMBC²⁹ allowed us to assign the imino protons from the assigned H8 protons of individual guanine residues (Figure 1.7a). Thus, H8 signals at δ_{H} 7.40 (G1), 8.02 (G11), 7.44 (G5) and 8.26 (G6) turned out to be correlated to the NH protons of their own base resonating at δ_{H} 12.00, 12.08, 12.23, 12.51, respectively. Furthermore, G8-H8 (δ_{H} 8.11) turned out to be correlated with the imino proton at δ_{H} 10.65. This low value for an imino proton suggests that G8 is not involved in classical Hoogsteen hydrogen-bonds.

Once the imino protons has been assigned, it is easy to determine the folding topology of sample **4**. In fact, by using the jump-return version of the 2D NOESY experiment, it was possible to detect a number of intense NOEs between the imino proton of a given base and the H8 proton of another base. Generally, this kind of NOE is very diagnostic in case the two bases adopt the Hoogsteen base pairing. In our case, we observed NOEs between G11-NH (12.08 ppm)/ G5-H8 (7.44 ppm) and G5-NH (12.23 ppm)/G2-H8 (8.04 ppm). These two correlations indicate that, as in TBA, G5 is involved in the formation of Hoogsteen hydrogen-bonds with both G11 and G2 and this is in agreement with the calculated structure (Figure 1.7b-c). We have also observed a NOE between G1-NH (12.00 ppm) and G6-H8 (8.26 ppm),

indicating that G1 and G6 are also paired. All of this indicates that the structure of **4** could be characterized also by a second G-triad formed by G10, G6 and G1. Unfortunately, due to unfavorable T1 noise, no NOE between G6-NH and G10-H8 could be unambiguously detected. Nevertheless, the fact that the subunit G10-G11 adopts a helical winding strongly supports the idea that actually also G10 does take part in the second G-triad.

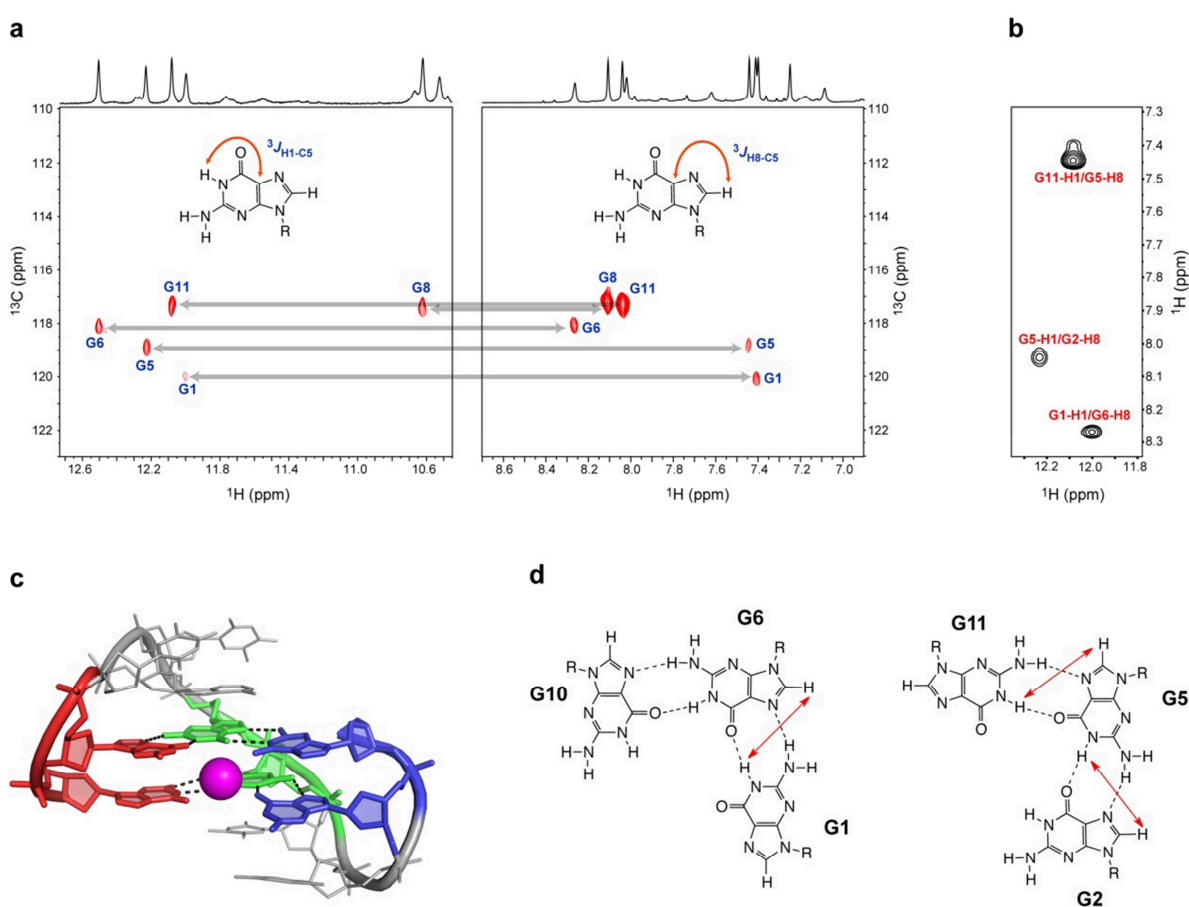


Figure 1.7: JR-HMBC and JR-NOESY spectra of 5'-GGTTGGTGTGG-3' (**4**). (a) Expanded regions of the JR-HMBC (600 MHz, T = 1 °C) showing correlations between H8 and H1 guanine protons with C5 carbons. (b) Expanded region of the 2D JR-NOESY (900 MHz, T = 1 °C) showing diagnostic correlations between H8 and H1 protons of paired bases. (c) Tridimensional representation of the G-triplex structure adopted by 5'-GGTTGGTGTGG-3' (**4**) as found in the MD calculations. (d) G-triads involved in the formation of the G-triplex. Red arrows show NOE correlations between H8 and H1 protons of paired bases.

In addition, the formation of the G-triplex is also inferred by the presence of a number of NOEs between NH protons and by the identification of other exchangeable protons. Particularly, in the 1D ^1H -NMR spectrum, along with the intense and narrow signals of the assigned imino protons (δ_{H} 12.00, 12.08, 12.23, 12.51) a number of other exchangeable protons could be observed. In order to clarify whether these signals could be attributable to other imino or amino protons, a 2D ^1H - ^{15}N HSQC experiment was acquired at 25 °C. At this temperature, the already assigned signals with δ_{H} (T = 1 °C) of 12.00, 12.08, 12.23, 12.51 and 10.65, were shifted respectively at δ_{H} 11.92, 11.99, 12.14, 12.37 and 10.45. These signals turned out to be correlated to nitrogens resonating at δ_{N} 143.37, 144.36, 144.66, 145.74 and 144.67 respectively. All these values are characteristic of N1 guanine bases. It is interesting to note that also the broad signal at δ_{H} 11.52 (25 °C) turned out to be correlated to similar frequencies (143.72 ppm), indicating that also this resonance is attributable to an imino proton of guanine. Unfortunately, no correlation could be observed in the JR-HMBC spectrum for this signal and it could not be unambiguously assigned to the pertinent base. Nevertheless, this signal is correlated with T9-H6,H1',H2'',Me; T7-Me; G10-H2'' and G8-H4'. All of this can be interpreted assuming that the imino proton resonating at δ_{H} 11.52 is assigned to G10. Probably, the signals of this NH proton is observable because involved in hydrogen bond. In fact, taking into account that the strand G14-G15 originally present in TBA is now missing, it is reasonable to assume that the strands G1-G2 and G10-G11 tend to bring themselves closer to each other, so that the formation of hydrogen bonds between G1 and G2 and G10 and G11, respectively, can not be ruled out. Furthermore, the formation of the G-triplex is also inferred by the

presence of a number of NOEs between NH protons. In particular, we observed strong NOEs between G1-NH/G5-NH and G6-NH/G11-NH; medium intensity NOEs between G1-NH/G6-NH and G5-NH/G11-NH; and weak NOEs between G5-NH/G6-NH, that once again support the formation of the G-triplex, having two G-triplets: G1-G6-G10 and G2-G5-G11 characterized by *syn-anti-syn* and *anti-syn-anti* arrangement of the residues, respectively.

Finally, a number of other NOEs definitely confirmed the structure of the G-triplex. Particularly, T7-Me shows correlations with G1-NH, G6-NH and G10-NH, confirming that the loop TGT is spatially very close to the G-triplet G1-G6-G10. Then, G11-NH shows correlations with T4-H2',H2'',H4',H5',H6 and T3-H4',H5',H5'',H6. Furthermore, G5-NH is also correlated with T3-H3', indicating that the TT loop is very close to the G-triplet G2-G5-G11.

MD simulation on the 11-mer oligonucleotide (4). The 11-mer oligonucleotide, **4**, in the G-triplex conformation has also been studied through an extensive MD calculation, in which we have reproduced the ionic concentration of 70 mM KCl used in the experiments. The G-triplex structure is very stable throughout the whole simulation (Figure 1.4) and fully agrees with the distances between atoms measured by the NMR experiments.

1.3 Conclusions

During the folding process of the G-quadruplex DNA aptamer TBA, we have observed the formation of a new DNA structural motif, the G-triplex. The existence

of this structure has been inferred in a 11 mer oligonucleotide, whose structural and thermodynamics properties were characterized by advanced computations, NMR, CD and DSC experiments. At variance with the already known triplex structures characterized by triads having standard Watson-Crick base pairings, G-triplex presents G:G:G triad planes stabilized by array of Hoogsteen-like hydrogen-bonds (Figure 1.1). This is the first time that DNA is found in the G-triplex conformation and this finding definitely supports the hypothesis of the existence of G-triplex structures during the G-quadruplex folding process^{30,31}.

This discovery is an important breakthrough that paves the way to new horizons in biology. In fact, guanine-rich regions, potentially able to form G-triplex structures, are very abundant in the genome and our study provides the molecular bases and the tools to investigate the presence of these structures in the genome, their biological role and the way to interact with them. Furthermore, our study raises new questions: are G-triplex structures present in the folding process of other DNA or RNA structures? Are there small molecules able to bind and interact with G-triplex? Are G-triplex structures useful for the design of new aptamers? These are only some of the issues that are to be addressed in the near future.

1.4 Materials and Methods

Metadynamics simulations. The starting conformation coordinates for TBA were obtained in its G-quadruplex conformation (PDB ID code 1qdf)²¹. The system was solvated using the TIP3P water model³² and neutralized adding Na⁺ ions with one of these ions placed at the center of the G-tetrad planes. All the simulations were carried out using periodic boundary conditions and Particle Mesh Ewald to

treat long range electrostatic. Before doing metadynamics simulations the system was equilibrated through 25 ns MD under NPT conditions at 1 atm and 300 K using the parmbsc0 parameters, which is a modified version of the Amber force field adapted for nucleic acids³³⁻³⁵. The Amber charges were applied to the DNA and waters atoms. The PLUMED plugin³⁶ was used to carry out metadynamics calculations with the NAMD code³⁷.

The estimation $F(s,t)$ at time t of the free-energy surfaces $F(s)$ as a function of the CVs was determined by metadynamics¹⁰ in its new well-tempered variant¹⁵, using the following formula:

$$F(s, t) = -\frac{T + \Delta T}{\Delta T} V(s, t) ,$$

where $V(s,t)$ is the bias potential added to the system and T is the temperature of the simulation. ΔT is the difference between the fictitious temperature of the CV and the temperature of the simulation. The bias potential is made up by the sum of the Gaussians deposited along the trajectories of the CVs. Thanks to this new formalism, one can increase barrier crossing and facilitate the exploration in the CVs space by tuning ΔT . A Gaussians deposition rate of 0.2 kcal/mol per picosecond was initially used and gradually decreased on the basis of the adaptive bias with a ΔT of 2,700 K. Given the good results previously published^{18,19}, we used the same CVs of their work. In particular, the first CV is the radius of gyration (R_{core}) calculated on the oxygen atoms of the guanines forming the G-tetrad planes. This collective variable is calculated using the following formula:

$$R_{core} = \left(\frac{\sum_i^n |r_i - r_{com}|^2}{\sum_i^n m_i} \right)^{1/2},$$

where the sums are over the n atoms and the center of mass is defined by

$$r_{com} = \frac{\sum_i^n r_i m_i}{\sum_i^n m_i}.$$

The gyration radius is a common descriptor in protein folding studies because it is able to discriminate between completely unfolded and globule states. A Gaussian width of 0.02 Å was used for this CV. In order to distinguish between globule and folded states, we used as second CV the number of intramolecular hydrogen-bonds engaged by the guanine bases forming the G-tetrad planes (H_{core}). The number of hydrogen-bonds is evaluated using the switching function

$$H_{core} = \sum_{ij} \frac{1 - \left(\frac{d_{ij}}{r_0}\right)^n}{1 - \left(\frac{d_{ij}}{r_0}\right)^m},$$

where r_0 is set to 2.0 Å, n and m are set to 6 and 12, respectively, i and j are the donor and acceptor hydrogen-bond atoms of the guanines used to calculate the number of hydrogen-bonds (Table 1.2).

On the latter CV we have used the reweighting algorithm²⁰ to compute the FES. In fact, this algorithm is able to reconstruct the Boltzmann distribution of CVs different from those used originally in the metadynamics run. Once the free energy

of the metadynamics simulation is converged, using the newly computed probability distribution, the reweighting method allows to build the FES on these CVs.

CV type	Atoms	
R_{core}	O6 (G1); O6 (G2); O6 (G5); O6 (G6); O6 (G10); O6 (G11); O6 (G14); O6 (G15)	
H_{core}	H-bond acceptor	H-bond donor
	O6 (G1)	H1 (G15)
	N7 (G1)	H21 (G15)
	O6 (G2)	H1 (G5)
	N7 (G2)	H21 (G5)
	O6 (G5)	H1 (G11)
	N7 (G5)	H21 (G11)
	O6 (G6)	H1 (G1)
	N7 (G6)	H21 (G1)
	O6 (G10)	H1 (G6)
	N7 (G10)	H21 (G6)
	O6 (G11)	H1 (G14)
	N7 (G11)	H21 (G14)
	O6 (G14)	H1 (G2)
	N7 (G14)	H21 (G2)
	O6 (G15)	H1 (G10)
	N7 (G15)	H21 (G10)

Table 1.2: List of the atoms used to define the R_{core} and H_{core} CV

NMR spectroscopy. All the experiments were collected on samples of the thrombin binding aptamer (TBA) and of its truncated constructs **1**, **2**, **3** and **4** at concentrations ranging from 0.7 mM up to 1.5 mM in potassium phosphate buffer (70 mM KCl, 10 mM KH₂PO₄, 0.2 mM EDTA, pH 7). NMR spectra were performed at 1 °C and 25 °C on Bruker spectrometers operating at 900, 700 and 600 MHz, equipped with triple resonance cryo-probes. Different pulse sequence schemes (presaturation, excitation sculpting^{38,39} and watergate⁴⁰) were used to suppress the water signal. 2D ¹H-¹H TOCSY and 2D ¹H-¹H COSY experiments were acquired at 700 MHz for the assignment of the spin systems. Proton–proton distance restraints were derived from the analysis of 2D ¹H-¹H NOESY and 2D ¹H-¹H ROESY acquired in H₂O at 900 and 800 MHz. 2D ¹H-¹H NOESY was acquired using several mixing times ranging between 50 and 300 ms. 2D ¹H-¹⁵N HSQC spectrum was acquired at 800 MHz to identify imino protons. Intranucleotide connectivities between imino and aromatic protons were obtained by a jump-and-return HMBC²⁹ spectrum acquired at 600 MHz and 950 MHz on an isotopically natural-abundance sample using 5120 transients. All spectra were processed with the Bruker TOPSPIN software packages and analyzed by the program CARA (Computer Aided Resonance Assignment, ETH Zurich)⁴¹.

CD and DSC experiments. Oligonucleotide samples were prepared by using the same buffer solution used for NMR experiments. CD spectra and CD melting curves of oligonucleotide samples were recorded by using a Jasco J-715 spectropolarimeter equipped with a Jasco JPT-423-S temperature controller using 1

mm path-length cuvettes. CD spectral scans were accumulated over the wavelength range 200–360 nm at 1 °C. The spectra were recorded at a scan rate of 100 nm/min with a response of 1 s, at 2.0 nm bandwidth and were averaged over 5 scans. Buffer baseline was subtracted from each spectrum. CD melting and annealing curves were recorded as a function of temperature in the range 0–90 °C at 289 nm with a scan rate of 0.5 °C/min. The CD melting curve of **4** was analyzed with a two-state model, using a theoretical equation for an intramolecular association, according to the van't Hoff analysis⁴². The T_m and $\Delta H^\circ_{\text{vH}}$ values provide the best fit of the experimental melting data.

Differential scanning calorimetry (DSC) measurements were carried out using a Nano DSC III (TA Instruments, New Castle, DE). The experiments were performed at oligonucleotide concentration of 0.2 mM. Scans were carried out at 0.5 and 1.0 °C/min scan rate in the temperature range 0–90 °C. Reversibility and repeatability were proven by multiple scans. A buffer-buffer scan was subtracted from the sample-buffer scans and linear-polynomial baselines were drawn for each scan. Baseline corrected thermograms were then normalized with respect to the oligonucleotide concentration to obtain the corresponding molar heat capacity curves. The model-independent transition enthalpies were obtained by integrating the area under the heat capacity versus temperature curves⁴². The melting temperatures were estimated as the temperatures corresponding to the maximum of each thermogram peak. Entropy values were obtained by integrating the $\Delta C_p/T$ versus T curves (where ΔC_p is the molar heat capacity and T is the temperature in kelvin) and the Gibbs energy values were computed by the equation $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$. The thermodynamic parameters reported are the averages of at least three

different heating experiments. The reported errors are the standard deviations of the mean from the multiple determinations.

References

- [1] Watson J. D. Crick, F., H. *Nature* **1935**, 171, 737-738.
- [2] Varani G. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, 24, 379–404.
- [3] Lilley D. M. J. *Proc. Natl Acad. Sci. USA* **1997**, 94, 9513-9515.
- [4] van de Sande J. H. *et al. Science* **1988**, 241, 551-557.
- [5] Sklená V. & Felgon J. *Nature* **1990**, 345, 836-838.
- [6] Parkinson G.N., Lee M. P. H., Neidle S. *Nature* **2002**, 417, 876-880.
- [7] Guéron M., Leroy J. L. *Curr. Opin. Struct. Biol.* **2000**, 10, 326-331.
- [8] Wells R. D. *Trends Biochem Sci.* **2007**, 32, 271-278.
- [9] Lipps H. J. and Rhodes D. *Trends in Cell Biology* **2009**, 19, 414-422.
- [10] Laio A., Parrinello M. *Proc. Natl. Acad. Sci., USA* **2002**, 99, 12562-12566.
- [11] Limongelli V. *et al. Proc. Natl. Acad. Sci. USA* **2012**, 109, 1467-1472.
- [12] Grazioso G. *et al. J. Am. Chem. Soc.* **2012**, 134, 453-463.
- [13] Limongelli V. *et al. Proc. Natl. Acad. Sci. USA* **2010**, 107, 5411-5416.
- [14] Rimón G. *et al. Proc. Natl. Acad. Sci. USA.* **2010**, 107, 28-33.
- [15] Barducci A., Bussi G., Parrinello M. *Phys. Rev. Lett.* **2008**, 100, 020603 .
- [16] Parrinello M. *eds. A. H. Zewail (Imperial College Press).* **2008**, pp 247-265.
- [17] Bock O. C., Griffin L. C., Latham J. A., Vermaas E. H. and Toole J. J. *Nature* **1992**, 355, 564-566.
- [18] Bussi G., Gervasio F. L., Laio A., Parrinello M. *J. Am. Chem. Soc.* **2006**, 128, 13435-13441.
- [19] Berteotti A., Barducci A., Parrinello M. *J. Am. Chem. Soc.* **2011**, 133, 17200-17206.
- [20] Bonomi M., Barducci A., Parrinello M. *J. Comput. Chem.* **2009**, 30, 1615-1621.
- [21] Marathias V. M. *et al. J. Mol. Biol.* **1996**, 260, 378-94.
- [22] Feigon J., Smith F. W. *Nature* **1992**, 356, 164-168.
- [23] Masiero S., Trotta R., Pieraccini S., De Tito S., Perone R., Randazzo A., Spada G. P. *Org Biomol Chem.* **2010**, 8, 2683-2692.
- [24] Karsisiotis A. I., Hessari N. M., Novellino E., Spada G. P., Randazzo A., Webba da Silva M. *Angew. Chem. Int. Ed. Engl.* **2011**, 50, 10645-10648.

- [25] Plum G. E., Breslauer K. J. *Curr. Opin. Struct. Biol.* **1995**, 5, 682-690.
- [26] Petraccone L., Pagano B., Esposito V., Randazzo A., Piccialli G., Barone G., Mattia C. A., Giancola C. *J. Am. Chem. Soc.* **2005**, 127, 16215-16223.
- [27] Smirnov I., Shafer R. H. *Biochemistry* **2000**, 39, 1462-1468.
- [28] Schultze P., Macaya R. F., Feigon, J. *J. Mol. Biol.* **1994**, 235, 1532-1547.
- [29] Phan A. T. *J. Biomol. NMR.* **2000**, 16, 175-178.
- [30] Mashimo T., Yagi H., Sannohe Y., Rajendran A., Sugiyama H. *J. Am. Chem. Soc.* **2010**, 312, 14910-14918.
- [31] Bončina M., Lah J., Prislán I., Vesnaver G. *J. Am. Chem. Soc.* **2012**, 134, 9657-9663.
- [32] Jorgensen W. L., Madura J. D. *J. Am. Chem. Soc.* **1983**, 105, 1407-1413.
- [33] Cornell W. D. *et al. J. Am. Chem. Soc.* **1995**, 117, 5179-5197.
- [34] Cheatham T. E. 3rd, Cieplak P., Kollman P. A. *J. Biomol. Struct. Dyn.* **1999**, 16, 845-862.
- [35] Pérez A. *et al. Biophys. J.* **2007**, 92, 3817-3829.
- [36] Bonomi M. *et al. Comp. Phys. Comm.* **2009**, 180, 1961-1972.
- [37] Phillips J. C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R. D., Kale L., Schulten, K. *J. Comput. Chem.* **2005**, 26, 1781-1802.
- [38] Hwang T. L., Shaka A. J. *J. Magn. Reson.* **1995**, A112, 275-279.
- [39] Dalvit C. *J. Biomol. NMR* **1998**, 11, 437-444.
- [40] Piotto M., Saudek V., Sklenar V. *J. J. Biomol. NMR* **1992**, 2, 661-665.
- [41] Keller R. (CANTINA Verlag, Goldau), **2004**.
- [42] Marky L. A., Breslauer, K. J. *Biopolymers* **1987**, 126, 1601-1620.

CHAPTER V

**PRELIMINARY STRUCTURE-FUNCTION
INVESTIGATIONS OF HsLARP4**

1) La protein

1.1 La protein: an RNA-binding protein

The La protein is a highly abundant nuclear phosphoprotein that often acts as an autoantigen in systemic lupus erythematosus and Sjogren's syndrome patients. Interestingly, La has been exhibited key roles in the processing and use of coding and noncoding RNAs^{1,2}. In the nucleus, La facilitates the production of tRNAs, acting as a RNA polymerase III (RNAP III) transcription factor by binding to the U-rich 3'UTR of nascent transcripts, assisting in their folding and maturation (e.g., pre-tRNA, 5S rRNA, and U6 and RNase P snRNAs)^{1,2,3}. It has been shown that the binding of La prevents degradation by 3' exonucleases^{4,5} but can even promote the nuclear retention of certain RNA precursors^{6,7}, and assist the pre-tRNA maturation^{8,9} and facilitate the proper assembly of other pol III transcripts, such as U6 RNA, into functional RNA:protein complexes⁵.

In yeast, La confirmed its function of stabilizing and protecting RNAs against exonucleolytic nibbling, binding to the 3' poly(U) ends that become exposed during endonucleolytic processing of pol II transcripts (e.g., snRNAs)^{5,10}.

Furthermore, La can directly facilitate translation of specific cellular^{11,12,13} and viral mRNAs^{14,15,16} usually by binding internal RNA sequences within the untranslated regions of target mRNAs. Thus, La exhibits remarkable versatility in its modes of binding to RNA.

Human La (HsLa) is a 408 amino acid monomeric modular protein containing three structured domains: a La motif and two RNA recognition motifs (RRMs 1 and

2); these are followed by a largely unstructured C-terminal tail that possesses a short basic motif (SBM) and a nuclear localization signal (NLS)^{1,2} (Figure 1.1).

Recently, in order to shed light on the molecular basis of La:RNA interactions, several studies have focused on the N-terminal domain (NTD), the most strongly conserved region of the protein between all eukaryotic species. The LaNTD is characterized by the presence of two domains: the La motif and RRM1, the domains required for high-affinity binding of 3' oligo(U) RNA^{17,18}; indeed, it has been shown that the two domains must act synergistically to bind 3' oligo(U) tails^{18,19,20}.

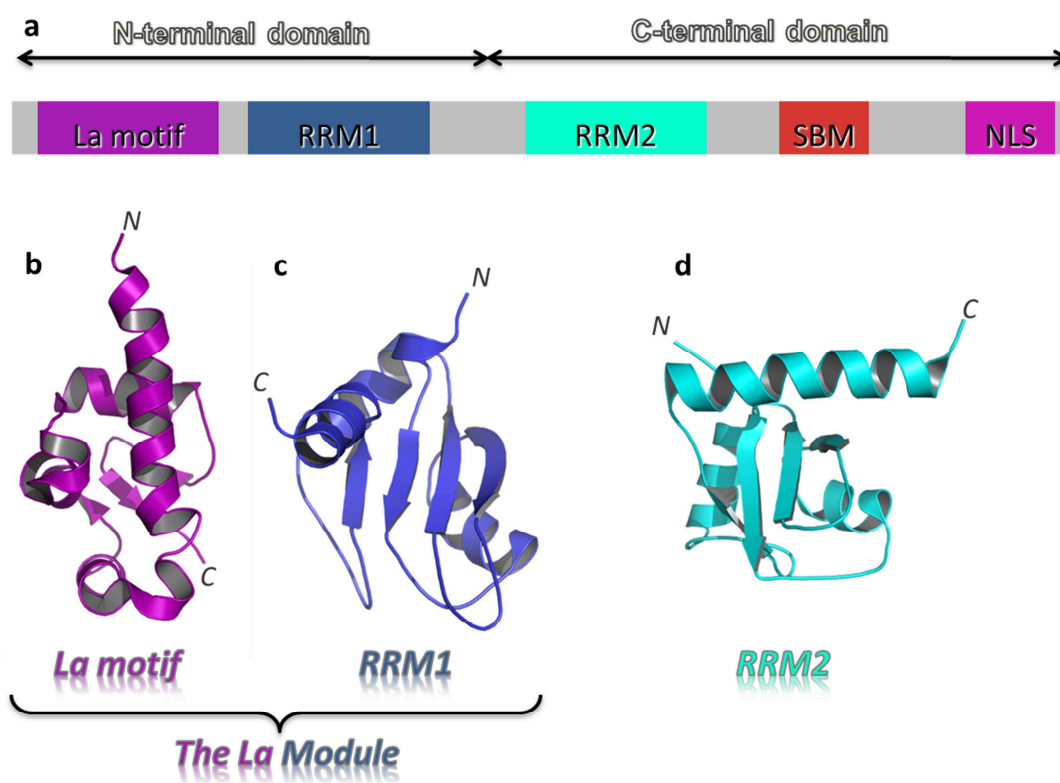


Figure 1.1: a) Domain organization of human La. Conventionally, human La is divided in two halves, N-terminal domain, comprising the La motif (b) and the adjacent RRM (c), and C-terminal domain, encompassing the RRM (d), the SBM and the NLS.

Structural studies on La have so far revealed several surprising aspects of the molecule. Although initially predicted to adopt an RRM fold, the controversial nature of the La motif domain^{1,2} was finally settled when it was shown to be an elaborated version of the winged-helix domain^{18,20}. Strikingly, chemical shift mapping experiments revealed that the recognition helix and wing loop of the La motif, which are normally used by this type of module to bind DNA²¹ or RNA²², are not involved in recognition of 3' oligo(U) sequences, indicating that La adopts a novel mode of interaction with RNA ligands¹⁸.

1.2 HsLa-RNA interaction

Previously it has been shown studies have established that both domains, the La motif and RRM1, were required for the interaction of 3'-UUU_{OH} with La^{17,23,24,25}; A detailed crystallographic study revealed that the recognition of a short RNA ending in 3'-UUU_{OH} by La domain from HsLa, involves other than to the expected surfaces²⁶, but nonetheless confirming the requirement for both motifs for RNA binding (Figure 1.2). Particularly, the structural analysis of this co-crystal has shown that the interaction of the protein with RNA induce the folding of the linker polypeptide connecting the two domains into an α helix. This appears to be an indirect effect since there are no contacts between the RNA and this portion of the polypeptide²⁶.

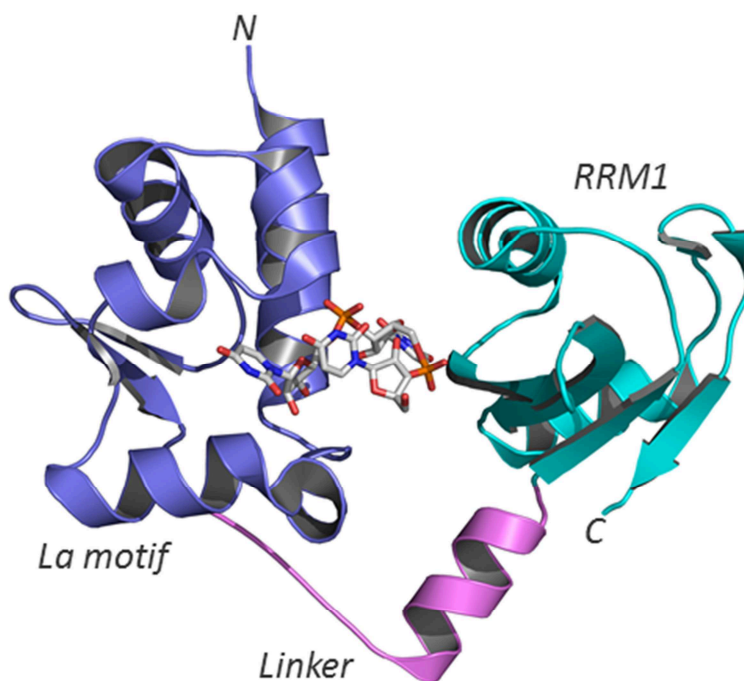


Figure 1.2: A co-crystal structure of binding modes of the La motif and RRM1 of the human La protein (NTD) with UUU RNA bound. La motif is in light blue, interdomain linker in pink and RRM1 in cyan.

To explain the binding mode of La protein Kotik-Kogan *et al.*²⁶ have performed NMR and crystal studies with different RNA oligomers ended with 3'-UUU_{OH}; in particular they studied the following oligos: AUUUU, AUAUUU, AUAUUUU, UUUUUUUU for the crystal structures, whilst for the NMR study they used UU, UUU, UUUU, UCUU, AUUUU, UUUUUUUU RNA oligomers.

The crystal structure of the complex LaNTD: AUUUU revealed a number of important structural details of the recognition of the 3' end of the RNA oligomer that help to explain the specificity of the LaNTD for sequences terminating in 3'-UUU_{OH}. The major interactions are made by the terminal pair of nucleotides (U-1/U-2), which consistently make the closest contacts with the protein and appear to provide the majority of the binding energy for the interaction (Figure 1.3).

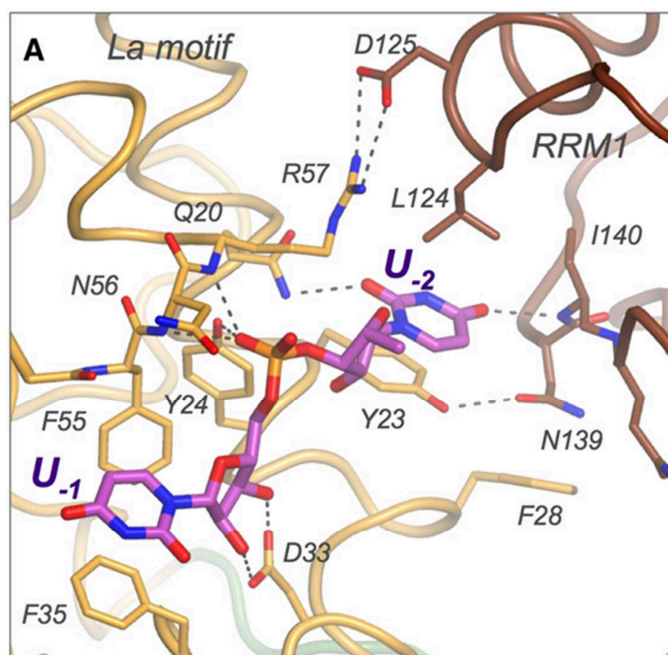


Figure 1.3: Structure of the LaNTD:AUUUU complex showing just the last two nucleotides, U-1 and U-2. The RNA is shown as a stick model. The protein is colored by domain: La motif orange and RRM1 brown; selected side chains are shown as sticks with their carbon atoms colored by domain. Hydrogen bonds are shown as dashed lines. (Taken from Kotik-Kogan *et al.* 2008²⁶)

The NMR studies have highlighted that the chemical shift perturbation result to be similar among all classes, revealing a common mode of 3' end recognition that involves both the La motif and RRM1 domains. In particular, all crystal structures revealed that the binding to RNAs appears largely based on the interactions established by the last two nucleotides in the sequence, U-1 and U-2. U-1 interacts exclusively with the La motif (Figure 1.3), interacting with specific H-bonds that involve only backbone features of the nucleotide: the 2' and 3'-OH groups from the ribose ring of U-1 are specifically recognized by hydrogen bonds to the O atoms of the side-chain carboxylate moiety of D33 (Figure 1.3), additionally, the O1 atom of the U-1 phosphate group makes hydrogen bonds with both the backbone NH groups of N56 and R57 and with the side-chain OH of Y24. Consistent with the

structure, these interactions with the ribose-phosphate backbone have been shown by mutagenesis to be important for 3' end recognition^{20,27}.

The penultimate nucleotide, U-2, makes intimate contacts with the protein that involve both the La motif and RRM1 (Figure 1.3). In contrast to U-1, it is the nucleotide base of U-2 and not its backbone that makes specific hydrogen bonds to the protein. Atoms O2 and O4 from the pyrimidine ring of U-2 hydrogen bond to the side-chain amide of Q20 (La motif) and the mainchain amide of I140 (RRM1), respectively, a pair of interactions that clearly helps to draw the two domains together. The pyrimidine ring of U-2 is also stacked directly on top of Y23, a residue from the La motif that is positioned by a hydrogen bond from the tyrosine side chain to the side chain of N139 of RRM1, which in turn stacks on top of F28 from the La motif (Figure 1.3). The U-2 pyrimidine ring is also packed directly underneath L124 from RRM1, a residue that is at least partly secured in place by the salt bridge made by its neighbour (D125) with R57 from the La motif. Thus, the recognition of U-2 in the complex engages a concerted set of protein:RNA and protein:protein interactions involving residues from both domains to form a tightly defined pocket that is specific in size, shape, and hydrogen-bonding capacity for a uridylate base; NMR analyses confirm that these interdomain interactions only form upon RNA binding. This induced fit of the binding pocket around U-2 accounts well for the cooperative nature of RNA binding by both domains of LaNTD^{17,18,19,20}. In the nucleus, La interacts mainly with 3'-oligoU containing RNAs, while in the cytoplasm La is able to recognize different mRNAs characterized by the absence of a 3'-UUU_{OH} trailer. An example of the latter is the binding of La to the IRES domain IV of the hepatitis C virus (HCV) RNA, which is involved in the stimulation of viral translation.

A biophysical investigation performed by Martino et. al, revealed that La binds to domain IV using a different binding mode from the one shown for its 3'-UUU_{OH} recognition: although the La motif and first RNA recognition motif (RRM1) are sufficient for high-affinity binding to 3'-oligoU, recognition of HCV domain IV requires not only the presence of the La motif and RRM1, but even the cooperation of the atypical RRM2, which to date was not involved in RNA binding. This new mode of binding does not depend from the sequence appear sequence specific, but recognizes different conformations of the RNA, in particular a double-stranded stem flanked by single stranded extensions. This study is the first step for a better understanding of the role of La in viral translation initiation.

References

- [1] Maraia R. J., Intine R. V., *Mol Cell Biol.* **2001**, 21, 367–379.
- [2] Wolin S.L., Cedervall T. *Annu Rev Biochem.* **2002**, 71, 375-403.
- [3] Stefano J. E. *Cell.* **1984** Jan;36,145-54.
- [4] Yoo C. J., Wolin S. L. *Cell.* **1997**, 89, 393-402.
- [5] Xue D., Robinson D. A., Pannone B. K., Yoo C. J., Wolin S. L. *EMBO J.* **2000**, 19, 1650-60.
- [6] Intine R. V., Dundr M., Misteli T., Maraia R. J. *Mol Cell.* **2002**, 9, 1113-23.
- [7] Intine R. V., Dundr M., Misteli T., Maraia R. J. *Mol Cell Biol.* **2007**, 27, 3303-12.
- [8] Van Horn D. J., Yoo C. J., Xue D., Shi H., Wolin S. L. *RNA.* **1997**, 1434-43.
- [9] Fan H, Goodier J. L., Chamberlain J. R., Engelke D. R., Maraia R. J. *Mol Cell Biol.* **1998**, 18, 3201-11.
- [10] Kufel J., Allmang C., Chanfreau G., Petfalski E., Lafontaine D. L., Tollervey D. *Mol Cell Biol.* **2000**, 20, 5415-24.
- [11] Holcik M., Korneluk R. G. *Mol Cell Biol.* **2000**, 20, 4648-57.
- [12] Trotta R., Vignudelli T., Candini O., Intine R. V., Pecorari L., Guerzoni C., Santilli G., Byrom M. W., Goldoni S., Ford L. P., Caligiuri M. A., Maraia R. J., Perrotti D., Calabretta B. *Cancer Cell.* **2003**, 3, 145-60.
- [13] Inada M., Guthrie C. *Proc Natl Acad Sci U S A.* **2004**, 101, 434-9.
- [14] Craig A. W., Svitkin Y. V., Lee H. S., Belsham G. J., Sonenberg N. *Mol Cell Biol.* **1997**, 17, 163-9.
- [15] Ali N., Pruijn G. J., Kenan D. J., Keene J. D., Siddiqui A. *J Biol Chem.* **2000**, 275, 27531-40.
- [16] Costa-Mattioli M., Svitkin Y., Sonenberg N. *Mol Cell Biol.* **2004**, 24, 6861-70.
- [17] Goodier J. L., Fan H., Maraia R. J. *Mol Cell Biol.* **1997**, 17, 5823-32.
- [18] Alfano C., Sanfelice D., Babon J., Kelly G., Jacks A., Curry S., Conte M. R. *Nat Struct Mol Biol.* **2004**, 11, 323-9.
- [19] Ohndorf U. M., Steegborn C., Knijff R., Sondermann P. *J Biol Chem.* **2001**, 276, 27188-96.
- [20] Dong G., Chakshusmathi G., Wolin S. L., Reinisch K. M. *EMBO J.* **2004**, 10, 1000-7.

- [21] Gajiwala K. S., Burley S. K. *Curr Opin Struct Biol.* **2000**, 10, 110-6.
- [22] Yoshizawa S., Rasubala L., Ose T, Kohda D., Fourmy D., Maenaka K. *Nat Struct Mol Biol.* **2005**, 12, 198-203.
- [23] Park J. M., Intine R. V., Maraia R. J. *Gene Expr.* **2007**,14, 71-81.
- [24] Martino L., Pennell S., Kelly G., Bui T. T., Kotik-Kogan O., Smerdon S. J., Drake A. F., Curry S., Conte M. R. *Nucleic Acids Res.* **2012**, 40,1381-94.
- [25] Horke S., Reumann K., Schulze C., Grosse F., Heise T. *J Biol Chem.* **2004**, 279, 50302-9.
- [26] Kotik-Kogan O., Valentine E. R., Sanfelice D., Conte M. R., Curry S. *Structure.* **2008**, 16, 852-62.
- [27] Teplova M., Yuan Y. R., Phan A. T., Malinina L., Ilin S., Teplov A, Patel D. J. *Mol Cell.* **2006**,21, 75-85.

2) La-Related Proteins

2.1 LARPs

The presence of the La Module (La motif plus RRM1) was even exhibited by a group of proteins otherwise unrelated to genuine La, defined as La-related proteins (LARPs) ¹⁻⁹. Functional and structural studies on LARPs are still in their infancy and so far only a few LARPs have been characterized. Through phylogenetic analyses and structural motif searches LARPs have been divided into four clearly distinct families (LARP1, 4, 6, and 7). The greatest degree of homology from the structural point of view between these various LARPs occurs in the La motif and also the RRM1 (Figure 2.1).

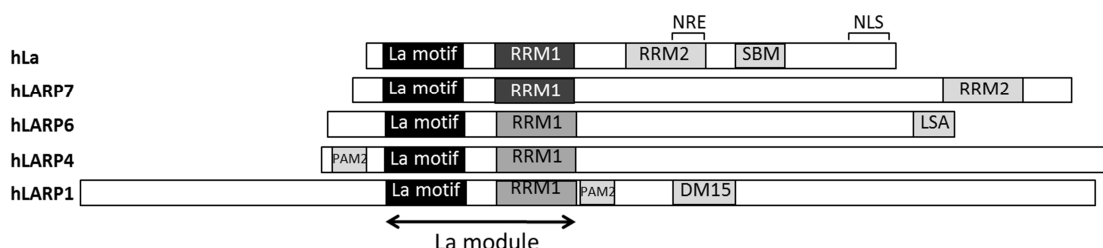


Figure 2.1: Schematic alignment indicating the conserved La motif-RRM domain architecture of the genuine La proteins followed by LARPs.

Initially, because of their similarity to La genuine protein, the researchers thought that LARPs also had same features of binding, in other words that they were capable to bind ligand with 3'-UUU_{OH} ended, but this has not always proved true. In fact, whereas some LARPs are indeed capable to recognize the RNA ending with the 3'-UUU_{OH} like the genuine La protein, LARP4 shows a different mechanism of recognition in that instead of binding poly (U) RNA it binds poly (A) RNA. In

humans two LARP4 family members are found (HsLARP4a and HsLARP4b, previously called HsLARP5)¹⁰, which share 37% identity and 53% similarity throughout their amino acid sequence. Unlike other LARPs, LARP4 and LARP4b diverge in several of the amino acids critical for canonical 3'-UUU_{OH} recognition, suggesting that, as already outlined, that the LARP4 family is the most diverged from genuine La proteins and other LARPs in the RNA binding pockets of their La Module¹¹.

2.2 Role of LARP4

The La motif-RRM arrangement has been found in La-related proteins 1 (LARP1), 1b, 4, 4b, 6, and 7, which have been separately conserved during evolution^{10,11,12} (LARP4b is also referred to as LARP5 in multiple databases and here will be designated LARP5/4b). LARP1, -6, and -7 have conserved all of the amino acids involved in 3'-UUU_{OH} recognition in La- RNA crystals^{13,14}, while LARP4 and -5/4b have diverged, suggesting alternative RNA binding¹¹. Moreover, an invariant divergence in all of the LARP4 and -5/4b sequences available occurs in a most critical residue involved in base-specific recognition seen in La-RNA crystals, corresponding to human La Q20, suggesting a conserved difference in RNA recognition¹¹. A detailed overview of the role and functions of LARP4 occurred in 2010 with a study conducted by Conte's group in collaboration with others¹⁵. According to their results, LARP4 appears to be involved in mRNA metabolism and translation. Multiple data indicate that LARP4 is intimately associated with translating polyadenylated mRNAs. The conserved RNA binding La motif-RRM of LARP4 exhibits a binding preference for poly(A), as well as other unexpected

properties of a La motif-RRM-containing protein. The interaction of LARP4 with RNA has been investigated by EMSA assays, where the binding properties of the La modules of human La(1-235) and LARP4(1-286) have been compared. Among the homopolymers tested, A(20) exhibited the best affinity for LARP4-NTD, as reflected by the ratio of free to bound RNA, whereas U(20) showed less binding and C(20) and G(20) showed no binding (Figure 2.2 A to D).

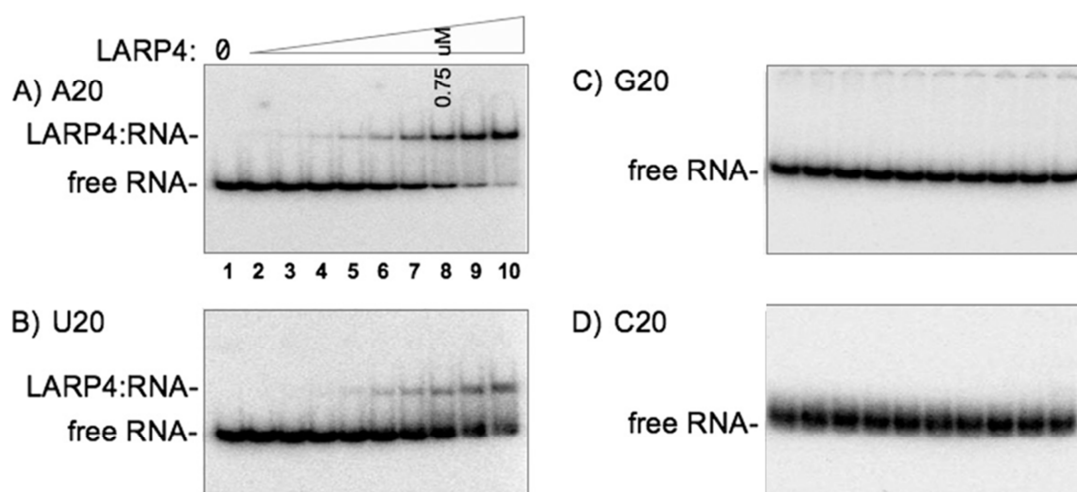


Figure 2.2: EMSA of homopolymeric 20-mer RNAs for binding to the RNA binding domain of LARP4 protein. (Taken from Yang R. *et al.*¹⁵)

LARP4 binding to RNA was not affected by addition of 3' phosphate; this is a difference with La protein that is capable to bind only ssRNA ending with 3' hydroxyl (Figure 2.3).

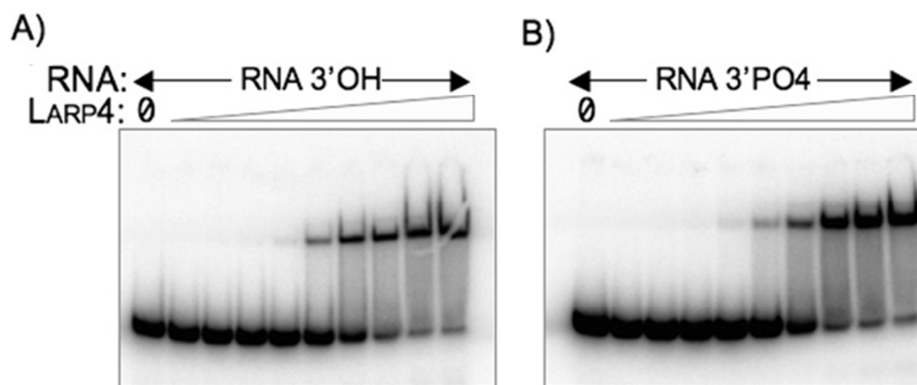


Figure 2.3: Comparison of LARP4(1-286) binding to otherwise identical 36-mer RNAs that end with a 3'-OH or a 3' phosphate. (Taken from Yang R. *et al.*¹⁵)

ITC (Figure 2.4) was employed to investigate in more detail the RNA binding properties of the La module of LARP4. The interaction of LARP4(111-303) with three different RNAs was tested.

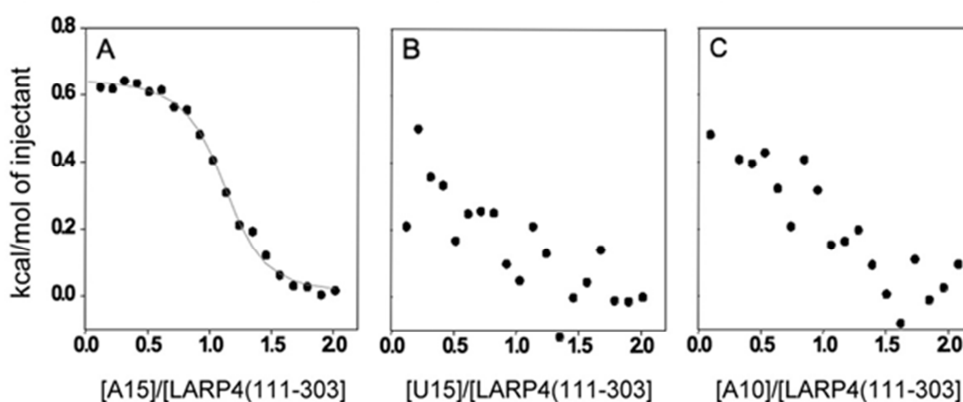


Figure 2.4: (A to C) Isothermal titration calorimetric analysis of LARP4(111-303) interactions with A(15), U(15), and A(10). The K_d and other thermodynamic parameters derived from this analysis are reported in Table 2.1. (Taken from Yang R. *et al.*¹⁵)

Figure 2.4 A to C shows the interactions of a LARP4(111-303) fragment with three RNAs. The titration analysis of LARP4(111-303) with A(15), has shown that binding occurs as one event centred on a molar ratio of 1. LARP4(111-303) interacts with A(15) with a K_d of 714 nM, with entropically driven binding (Table 2.1).

Protein	<i>n</i>	K_d (1/ K_b) (nM)	ΔH° (kcal mol ⁻¹)	$-T\Delta S^\circ$ (kcal mol ⁻¹)	ΔG°_{298K} (kcal mol ⁻¹)
LARP4(111-303)/U(15)	1.0	714	0.7 ± 0.1	-9.1 ± 0.3	-8.4 ± 0.4
LARP4(111-303)/A(10)					
LARP4(111-303) A(15)					

Table 2.1: Thermodynamic parameters of the interaction of La(1-194) and LARP4(111-303) with single-stranded RNA sequence. The values represent the averages and the standard deviations over three independent measurements

In contrast, LARP4(111-303) displayed very weak association with U(15) and A(10) with a K_d beyond the threshold that could be rigorously measured by ITC, *i.e.*, ≥ 0.1 mM. Therefore, LARP4 binds to A(15) at least 200-fold more tightly than U(15) and A(10).

The LARP4 La motif-RRM is flanked on both sides by PABP interacting regions. The region N-terminal to the La motif-RRM contains a conserved PAM2w sequence that we showed is important for PABP interaction. PAM2 sequences mediate direct contacts with the MLLE domain of PABP^{16,17}. Direct binding as monitored by ITC, as well as NMR and crystallography, showed that the LARP4 PAM2w makes contacts with the MLLE domain of PABP.

Notably, it has been shown evidence that is consistent with a model in which LARP4 and PABP interact in the presence of RNA and that this contributes to the stability of their interaction. This characterization represents the first detailed mechanism by which a LARP interacts with PABP.

References

- [1] Yoo C. J., Wolin S. L. *Mol Cell Biol.* **1994**, 14, 5412-24.
- [2] Sobel S. G., Wolin S. L. *Mol Biol Cell.* **1999**, 10, 3849-62.
- [3] Remillieux-Leschelle N., Santamaria P., Randsholt N. B. *Genetics.* **2002**, 162, 1259-74.
- [4] Aigner S., Postberg J., Lipps H. J., Cech T. R. *Biochemistry.* **2003**, 42, 5736-47.
- [5] Witkin K. L., Collins K. *Genes Dev.* **2004**, 18, 1107-18.
- [6] He N., Jahchan N. S., Hong E., Li Q., Bayfield M. A., Maraia R. J., Luo K., Zhou Q. *Mol Cell.* **2008**, 29, 588-99.
- [7] Krueger B. J., Jeronimo C., Roy B. B., Bouchard A., Barrandon C., Byers S. A., Searcey C. E., Cooper J. J., Bensaude O., Cohen E. A., Coulombe B., Price D. H. *Nucleic Acids Res.* **2008**, 36, 2219-29
- [8] Markert A., Grimm M., Martinez J., Wiesner J., Meyerhans A., Meyuhas O., Sickmann A., Fischer U. *EMBO Rep.* **2008**, 9, 569-75.
- [9] Nykamp K., Lee M. H., Kimble J. *RNA.* **2008**, 14, 1378-89.
- [10] Bousquet-Antonelli C., Deragon J. M. *RNA.* **2009**, 15, 750-64.
- [11] Bayfield M. A., Yang R., Maraia R. J. *Biochim Biophys Acta.* **2010**, 1799, 365-78.
- [12] Bayfield M. A., Yang R., Maraia R. J. *Dev Biol.* **2009**, 334, 186-97.
- [13] Kotik-Kogan O., Valentine E. R., Sanfelice D., Conte M. R., Curry S. *Structure.* **2008**, 16, 852-62.
- [14] Teplova M., Yuan Y. R., Phan A. T., Malinina L., Ilin S., Teplov A., Patel D. J. *Mol Cell.* **2006**, 21, 75-85.
- [15] Yang R., Gaidamakov S. A., Xie J., Lee J., Martino L., Kozlov G., Crawford A. K., Russo A. N., Conte M. R., Gehring K., Maraia R. J. *Mol Cell Biol.* **2011**, 31, 542-56.
- [16] Kozlov G., De Crescenzo G., Lim N. S., Siddiqui N., Fantus D., Kahvejian A., Trempe J. F., Elias D., Ekiel I., Sonenberg N., O'Connor-McCourt M., Gehring K. *EMBO J.* **2004**, 23, 272-81.
- [17] Kozlov G., Ménade M., Rosenauer A., Nguyen L., Gehring K. *J Mol Biol.* **2010**, 397, 397-407.

3) Study of the interaction between Human La-Related Protein 4 (HsLARP4) and poly-A15 RNA

3.1 Introduction

Human La-Related Protein 4 (HsLARP4) interacts with high affinity and specificity to poly-A15 RNA (see section 2.2). The La module of HsLARP4 is essential for the interaction with poly-A15 RNA.

In this project, five highly conserved residues in the La module of HsLARP4 were identified through sequence alignment with HsLa, and three of them were mutated to Alanine through point mutation (Figure 3.1). The aim of the study was to investigate if the highly conserved residues that play a key role in the interaction HsLa/3'-UUU_{OH}, result to be important also in the HsLARP4/poly-A15 interaction .

The mutants were used in Isothermal Titration Calorimetry (ITC) to investigate the interaction between HsLARP4 and poly-A15 RNA. Mutation of residues 126, 139 and 160 to Alanine in the La module of HsLARP4, decreased its binding affinity to poly-A15, hence, indicating that these residues of HsLARP4 play key roles in the interaction with RNA.

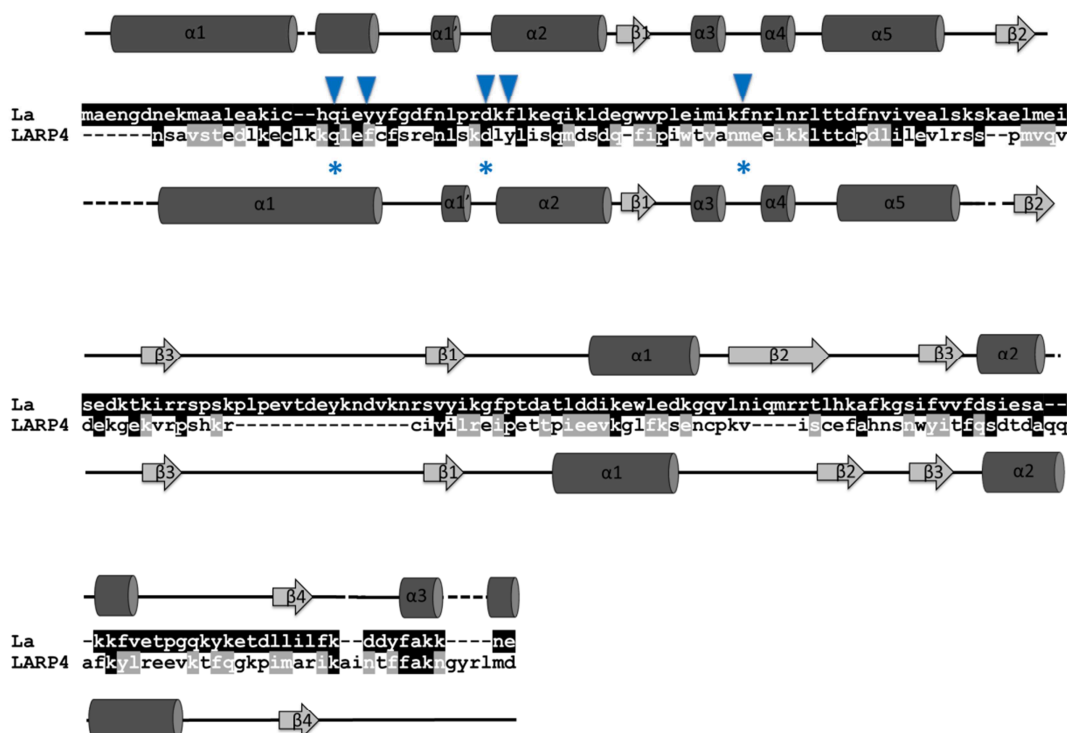


Figure 3.1: Sequence alignment of the conserved La module of HsLa and HsLARP4a. The residues that in HsLa are responsible for the 3' oligoU specific binding are indicated with blue arrowheads. The residues that were mutated into Alanine in HsLARP4 La module are indicated with blue stars. The secondary structure of HsLARP4 La motif is reported according to the NMR solution structure, while the secondary structure of the RRM1 domain has been derived from preliminary NMR data obtained with TALOS+.

3.2 Results and Discussion

Generation of plasmids encoding for three point mutations of HsLARP4(111-287). Contrary to the interaction of HsLa with poly U RNA, HsLARP4's interaction with poly A15 RNA was shown to be length-dependent (see section 2.2). In order to evaluate if the conserved residues in the La motif of the LARPs play a key role in the interaction of LARP4 with poly-A15 RNA, we selected three conserved residues (Q126, D139, M160) and mutate them into Alanine residues. This goal was achieved through the Polymerase Chain Reaction (PCR) using primers in Figure 3.37

(section 3.4). Agarose gel electrophoresis analysis showed bands of about 6 kilobases and smears for all three PCR products (Figure 3.2), and they are likely to be formed by the plasmids (5.1 kilobases) supercoiled in various conformations. Further, DNA sequencing results showed that the desired mutation was present in all plasmids.

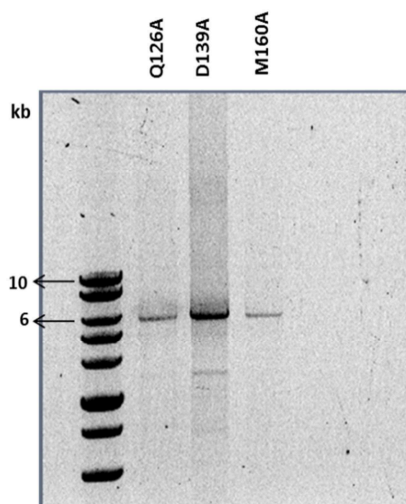


Figure 3.2. Agarose gel electrophoresis of the PCR products generated from the point mutation. First lane: 1Kb DNA ladder (Promega, USA), lane Q126A: Plasmid encoding for Q126A mutant, lane D139A: Plasmid encoding for D139A mutant, lane M160A: Plasmid encoding for M160A mutant.

Protein expression test in Rosetta2 cells at different temperature. The plasmids of section 3.4 were transformed into Rosetta2 cells, and expression test was performed at 18 °C and 37 °C. SDS-PAGE analysis in Figures 3.4 and 3.5 showed that there were over-expression of proteins around the 23 kDa mark in HsLARP4(111-287) wild-type and all mutants. The molecular weight of wild-type HsLARP4(111-287), Histidine-Tag (His-Tag) and Tobacco-Etch-Virus (TEV) cleavage site (see Figure 3.3) was about 23.1 kDa.

	MG SSHHHHHHSQ DPENLYFQSV				
111	NSAVSTEDLK	ECLKKQLEFC	FSRENLSKDL	YLISQMDSQ	150
151	FIPIWTVANM	EEIKKLTTDP	DLILEVLRSS	PMVQVDEKGE	190
191	KVRPSHKRCI	VILREIPETT	PIEEVKGLFK	SENC PKVISC	230
231	EFAHNSNWYI	TFQSDTDAQQ	AFKYLREEVK	TFQGKPIMAR	270
271	IKAINTFFAK	NGYRLMD ²⁸⁷			

Figure 3.3: Sequence of HsLARP4(111-287), Histidine-Tag (His-Tag) and Tobacco-Etch-Virus (TEV) cleavage site. The His-Tag and TEV site are reported in red, while the point mutations Q126A, D139A and M160A are reported in green.

Since in the point mutations only one residue of the wild-type sequence was replaced with an Alanine, all the mutants were expected to have similar molecular weight. More proteins were present in the soluble fraction when induced at 18 °C than 37 °C (data not shown).

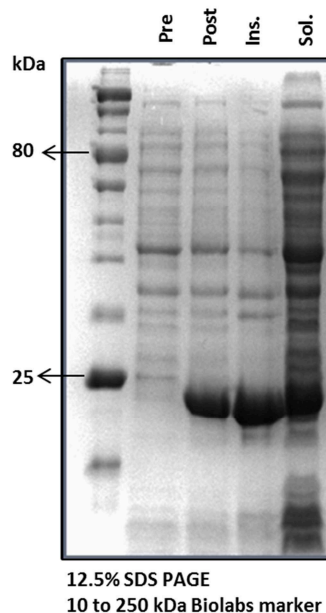


Figure 3.4. SDS-PAGE analysis of the expression test with induction done at 18°C on Rosetta2 cells transformed with plasmid encoding for HsLARP4(111-287). First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane Pre: Wild-type pre-induction sample, lane Post: post-induction sample, lane Ins.: insoluble fraction, lane Sol.: soluble fraction respectively.

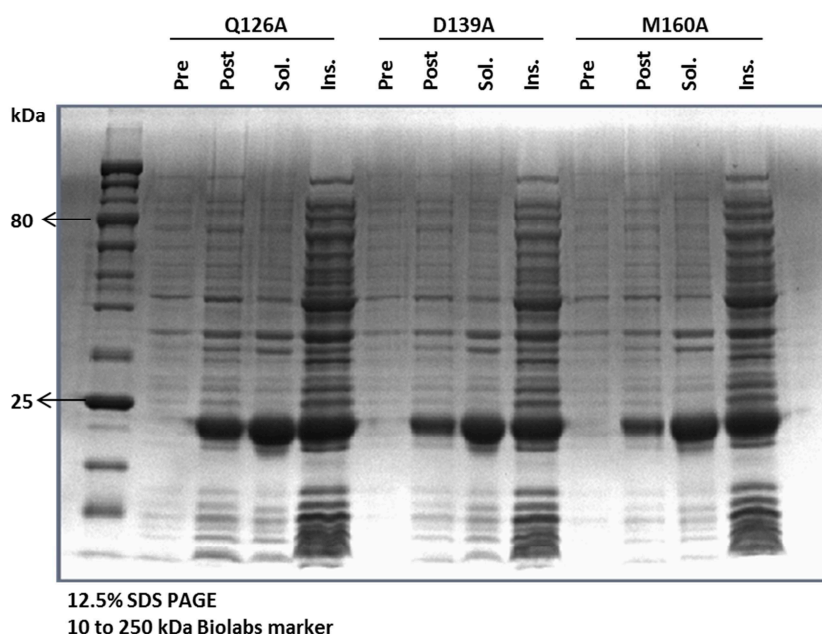


Figure 3.5. SDS-PAGE analysis of the expression test with induction done at 18°C on Rosetta2 cells transformed with plasmid encoding for HsLARP4(111-287) Q126A, HsLARP4(111-287) D139A a HsLARP4(111-287) M160A. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lanes Pre, Post, Sol., Ins. (Q126A): HsLARP4(111-287) Q126A pre-induction sample, post-induction sample, soluble fraction, insoluble fraction respectively, lanes Pre, Post, Sol., Ins. (Q126A): HsLARP4(111-287) D139A pre-induction sample, post-induction sample, soluble fraction, insoluble fraction respectively, lanes Pre, Post, Sol., Ins. (M126A): HsLARP4(111-287) M160A pre-induction sample, post-induction sample, soluble fraction, insoluble fraction respectively.

Isolation and purification of the HsLARP4(111-287) & the mutants.

Expression test has shown the over-expression of HsLARP4(111-287) wild-type and the three mutants in Rosetta2 cells on induction. Large-scale production, isolation and purification were done to produce sufficient proteins for NMR and ITC experiments. Purification of HsLARP4(111-287) wild-type, mutant Q126A, D139A and M160A were successful as reported in the following session.

HsLARP4(111-287). Figure 3.6 shows the elution profile of HsLARP4(111-287) wild-type from Nickel Column, with the peak occurring from fraction A9 to C1. SDS-PAGE analysis (Figure 3.7) confirmed the presence of the protein from fraction A12 to B9. Manual purification through Nickel column after TEV cleavage removed the His-tag, TEV protease and contaminating protein. In Figure 3.8, the size of the

protein reduced from about 23.1 kDa to 20.5 kDa after TEV cleavage. Purification through Heparin columns removed nucleic acids contaminants. Figure 3.9 shows elution profile from Heparin Column, with the peak occurring from fraction C4 to D3. SDS-PAGE analysis (Figure 3.10) confirmed the presence of the protein from fraction C4 to D3. Finally, in order to eliminate last residues of contaminants, size exclusion column (SEC) purification was performed. Figure 3.11 shows the elution profile from SEC. Two peaks appear in the chromatogram: one from fraction C9 to D5 and the other one from fraction E4 to F2. According to SDS-PAGE analysis (Figure 3.12), only the fractions from E4 to F2 were collected.

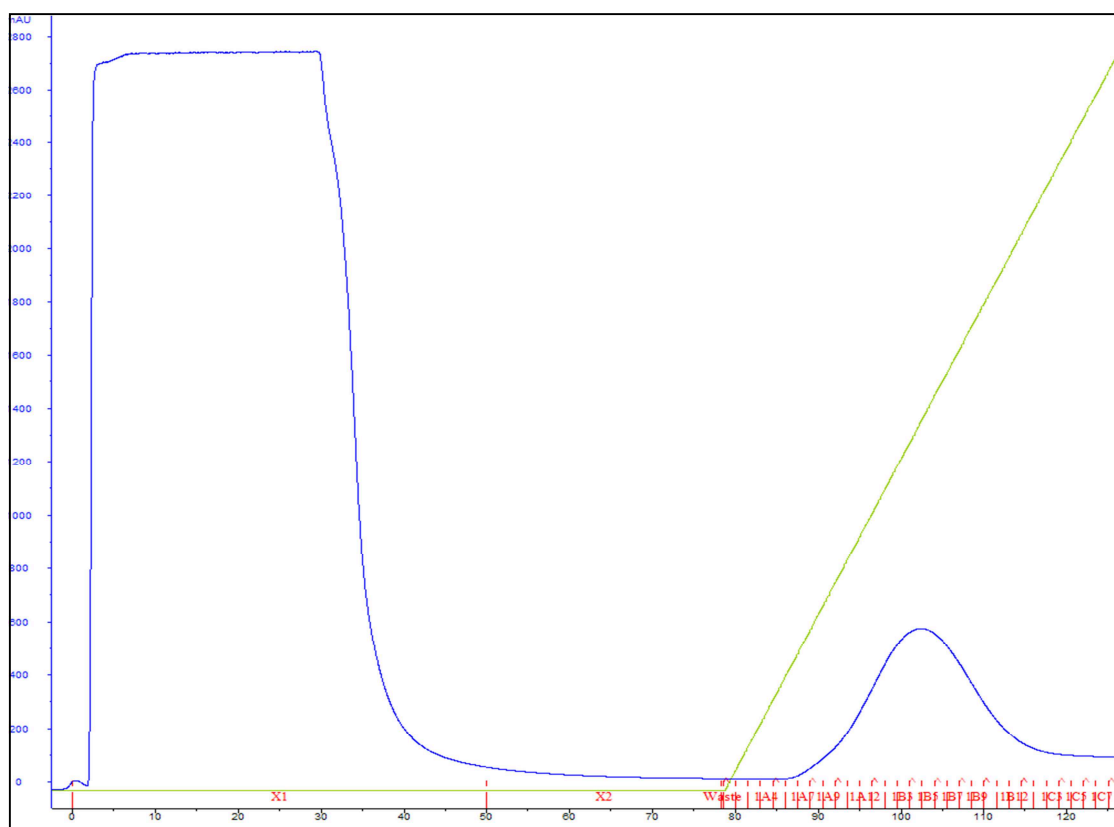


Figure 3.6: Elution profile HsLARP4(111-287) wild-type from Nickel affinity column using a gradient of Nickel buffer B following a wash with Nickel buffer A. Protein elution was monitored by UV absorbance at 280nm. Distinct peak of UV absorbance was observed from fraction A9 to C1. Blue curve: Ultraviolet Absorbance; Green curve: Nickel Buffer B concentration.

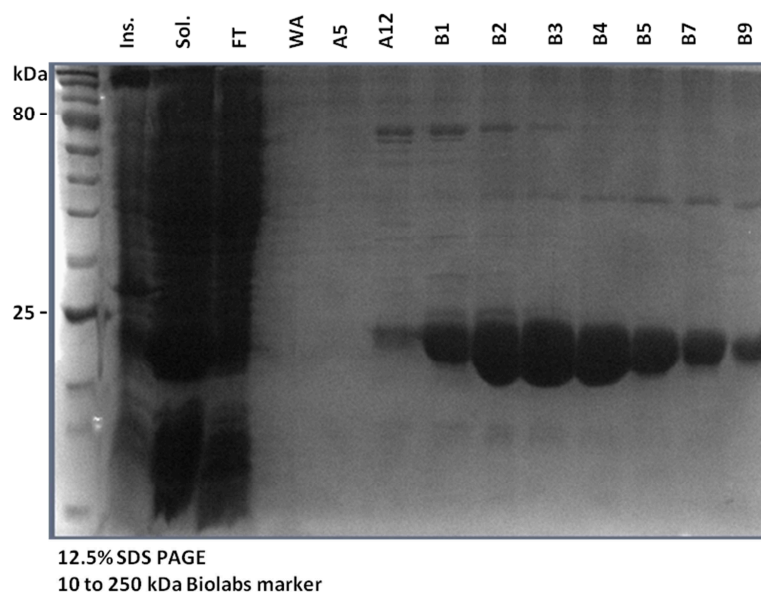


Figure 3.7: SDS PAGE analysis of fraction A5 to B9 from the Nickel column, showing that HsLARP4(111-287), of about 23.1 kDa, was found predominantly from fraction A12 to B9. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane Ins.: Insoluble cell fraction after sonication and centrifugation, lane Sol.: Soluble cell fraction after sonication and centrifugation lane FT: Flow-through from Nickel column, lane 5: Fraction collected after washing the column with Nickel buffer A, lane A5: Fraction A5, lane A12: Fraction A12, lane B1: Fraction B1, lane B2: Fraction B2, lane B3: Fraction B3, lane B4: Fraction B4, lane B5: Fraction B5, lane B7: Fraction B7, lane B9: Fraction B9.

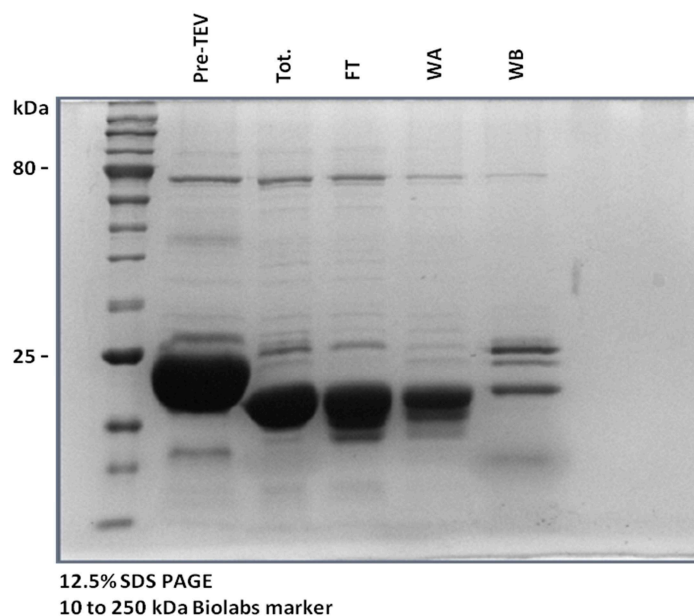


Figure 3.8: SDS-PAGE analysis of the protein solution after the 2nd round of purification through the Nickel affinity column after the removal of His-tag by TEV protease. Prior to TEV cleavage, the protein was about 23.1 kDa. After TEV cleavage, the molecular weight of the protein was reduced to about 20.5 kDa. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane Pre-TEV: protein solution prior to cleavage, lane Tot.: protein solution after cleavage, but prior to purification through the Nickel column, lane FT: protein solution after passing through the Nickel column and eluted in the flow-through, lane WA: Elution of the contaminating protein with Nickel Buffer A.

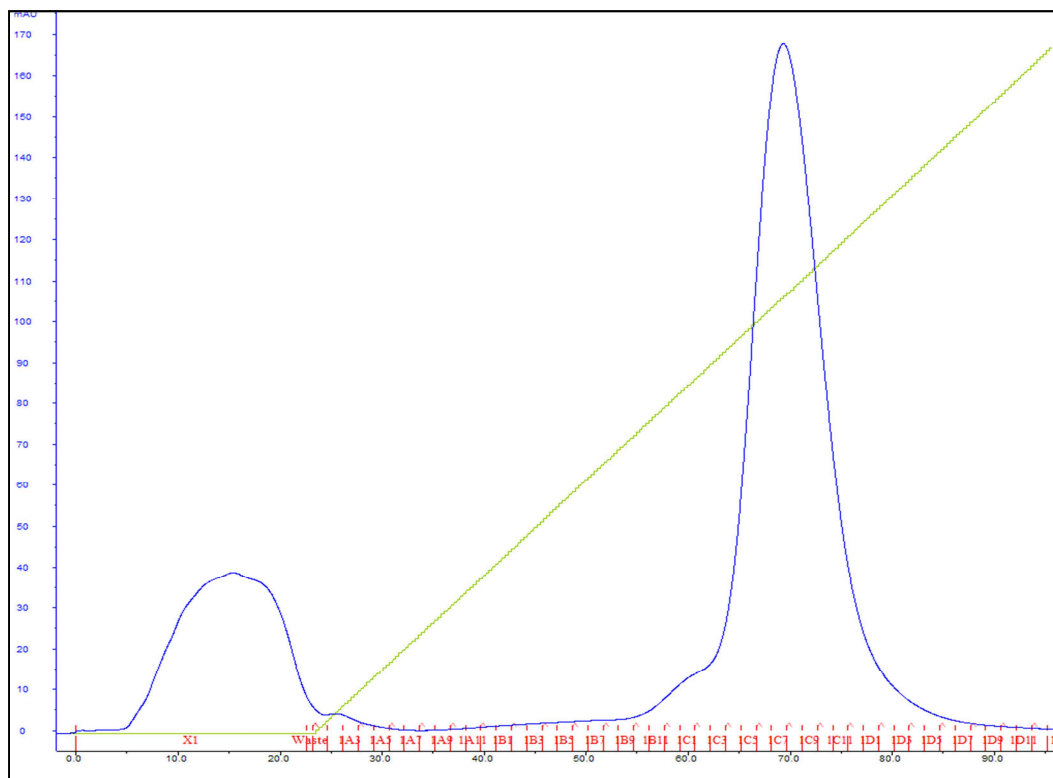


Figure 3.9: Elution profile HsLARP4(111-287) from the Heparin columns washed with Heparin Buffer A, and eluted with a Heparin Buffer B gradient. Protein elution was monitored by UV absorbance at 280nm. Distinct peak of UV absorbance was formed from fraction C4 to fraction D3. Blue curve: Ultraviolet Absorbance; Green curve: Heparin Buffer B concentration.

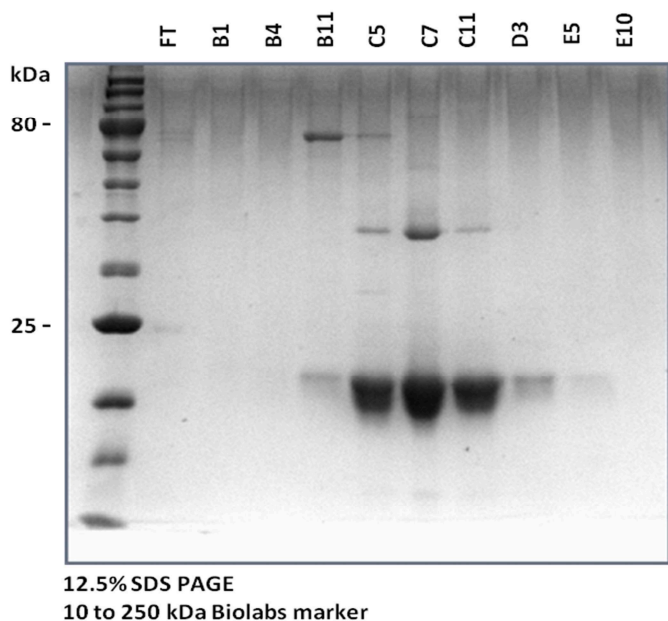


Figure 3.10: SDS PAGE analysis of fraction B1 to F10 from the Heparin columns, showing that HsLARP4(111-287) of about 20.5 kDa after removal of HisTag by TEV protease, was found predominantly from fraction E5 to E12. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane FT: Flow-through of protein solution after passing through the double Heparin columns, lane B1: Fraction B1, lane B4: Fraction B4, lane B11: Fraction B11, lane C5: Fraction C5, lane C7: Fraction C7, lane C11: Fraction C11, lane D3: Fraction D3, lane E5: Fraction E5, lane E10: Fraction E10.

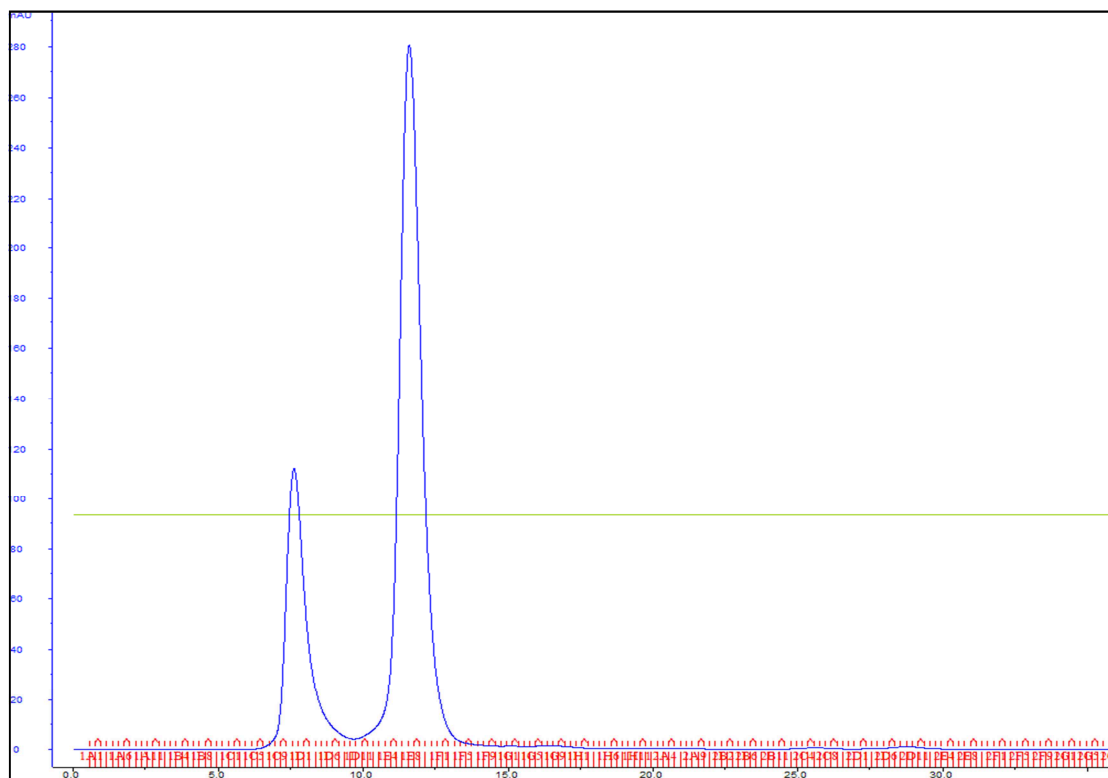


Figure 3.11: Elution profile HsLARP4(111-287) from the Size Exclusion column washed with NMR/ITC buffer. Protein elution was monitored by UV absorbance at 280nm. Distinct peaks of UV absorbance were visible from fraction C9 to fraction D5 and E4 to F2. Blue curve: Ultraviolet Absorbance; Green curve: NMR/ITC buffer concentration.

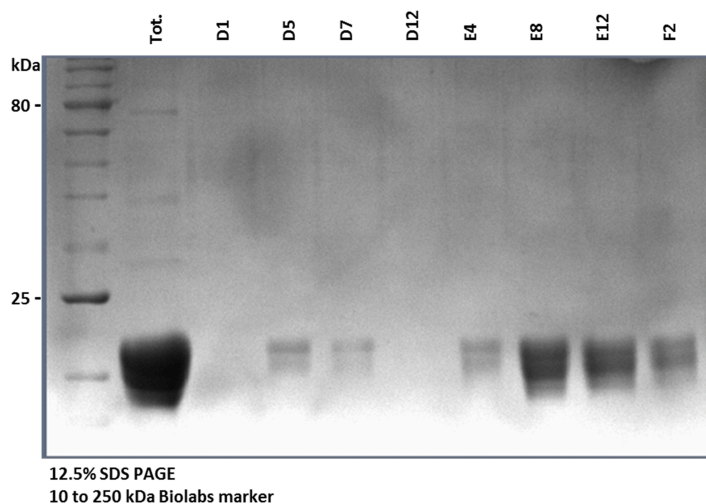


Figure 3.12: SDS PAGE analysis of fraction D1 to F2 from the size exclusion column, showing that HsLARP4(111-287) of about 20.5 kDa after removal of HisTag by TEV protease, was found predominantly from fraction E4 to F9. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane Tot.: Sample prior 3rd round of purification, lane D1: Fraction D1, lane D5: Fraction D5, lane D7: Fraction D7, lane D12: Fraction D12, lane E4: Fraction E4, lane E8: Fraction E8, lane E12: Fraction E12, lane F2: Fraction F2.

HsLARP4(111-287) Q126A. Figure 3.13 shows the elution profile of *HsLARP4(111-287) Q126A* from Nickel Column, with the peak occurring from fraction A9 to B12. SDS-PAGE analysis (Figure 3.14) confirmed the presence of the protein from fraction A11 to B12.

Manual purification through Nickel column after TEV cleavage removed the His-tag, TEV protease and contaminating protein. In Figure 3.15, the size of the protein reduced from about 23.1 kDa to 20.5 kDa after TEV cleavage.

Purification through Heparin columns removed nucleic acids contaminants. Figure 3.16 shows elution profile from Heparin Column, with the peak occurring from fraction C1 to D5. SDS-PAGE analysis (Figure 3.17) showed the protein presence from fraction C1 to D1.

Finally, in order to eliminate last residues of contaminants, size exclusion column (SEC) purification was performed. Figure 3.18 shows the elution profile from SEC, with a peak occurring from fraction E5 to F5. According to SDS-PAGE analysis (Figure 3.19), only the fractions from E5 to F5 were collected.

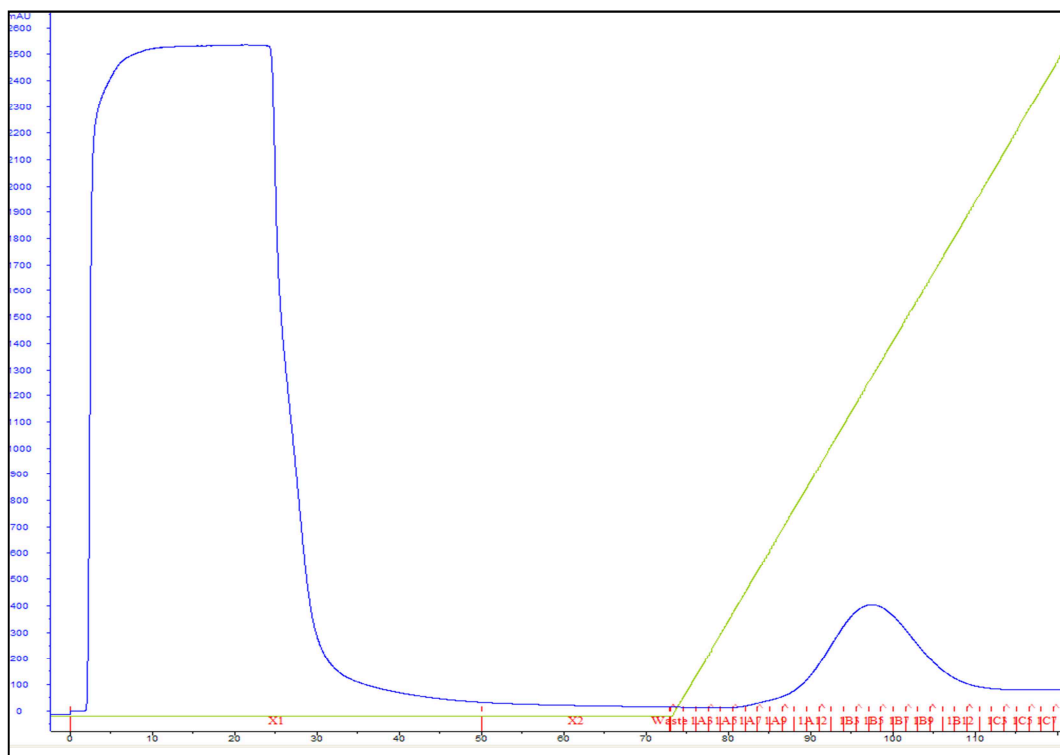


Figure 3.13: Elution profile HsLARP4(111-287) Q126A from Nickel affinity column using a gradient of Nickel buffer B following a wash with Nickel buffer A. Protein elution was monitored by UV absorbance at 280nm. Distinct peak of UV absorbance was observed from fraction A9 to B12. Blue curve: Ultraviolet Absorbance; Green curve: Nickel Buffer B concentration.

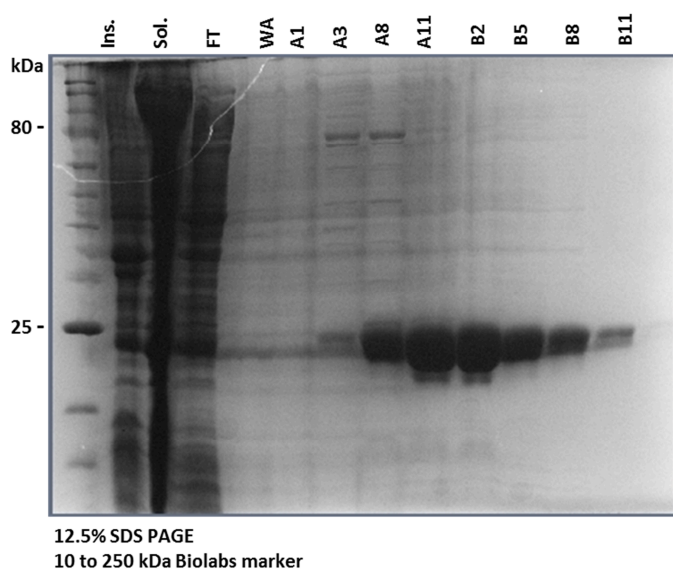


Figure 3.14: SDS PAGE analysis of fraction A1 to B11 from the Nickel column, showing that HsLARP4(111-287) Q126A, of about 23.1 kDa, was found predominantly from fraction A9 to B12. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane Ins.: Insoluble cell fraction after sonication and centrifugation, lane Sol.: Soluble cell fraction after sonication and centrifugation lane FT: Flow-through from Nickel column, lane 5: Fraction collected after washing the column with Nickel buffer A, lane A1: Fraction A1, lane A3: Fraction A3, lane A8: Fraction A8, lane A11: Fraction A11, lane B2: Fraction B2, lane B5: Fraction B5, lane B8: Fraction B8, lane B11: Fraction B11.

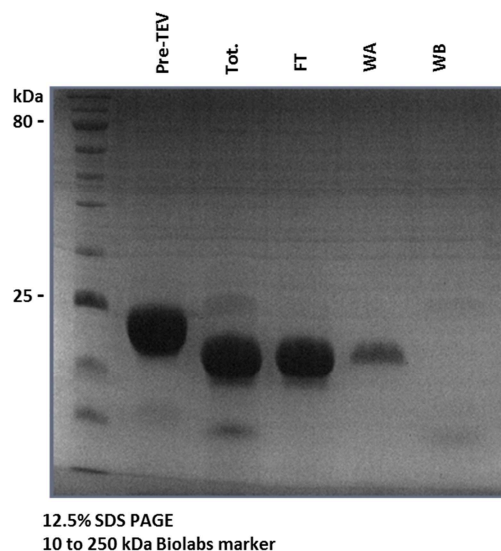


Figure 3.15: SDS-PAGE analysis of the protein solution after the 2nd round of purification through the Nickel affinity column after the removal of His-tag by TEV protease. Prior to TEV cleavage, the protein was about 23.1 kDa. After TEV cleavage, the molecular weight of the protein was reduced to about 20.5 kDa. lane 1: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane 2: protein solution prior to cleavage, lane 3: protein solution after cleavage, but prior to passing through the Nickel column, lane 4: protein solution after passing through the Nickel column and eluted out in the flow-through, lane WA: Elution of the contaminating protein by washing the column with Nickel Buffer A, lane WB: Elution of the contaminating protein by washing the column with Nickel Buffer B.

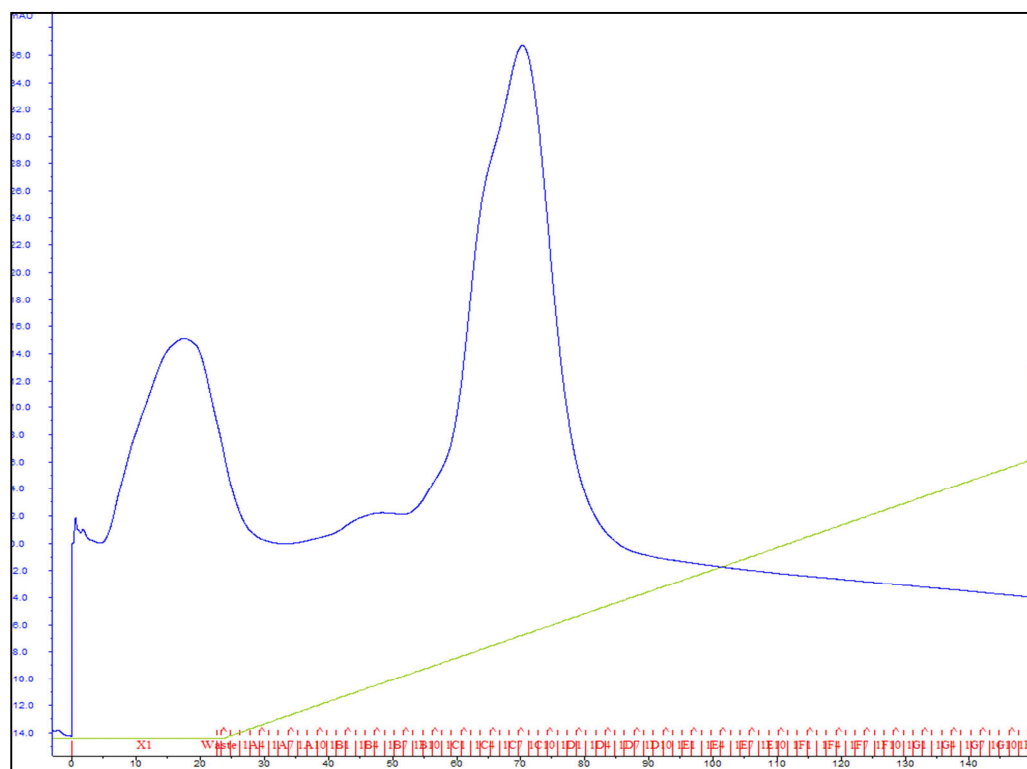


Figure 3.16: Elution profile HsLARP4(111-287) Q126A from the Heparin columns washed with Heparin Buffer A, and eluted with a Heparin Buffer B gradient. Protein elution was monitored by UV absorbance at 280nm. Distinct peak of UV absorbance was formed from fraction C1 to fraction D5. Blue curve: Ultraviolet Absorbance; Green curve: Heparin Buffer B concentration.

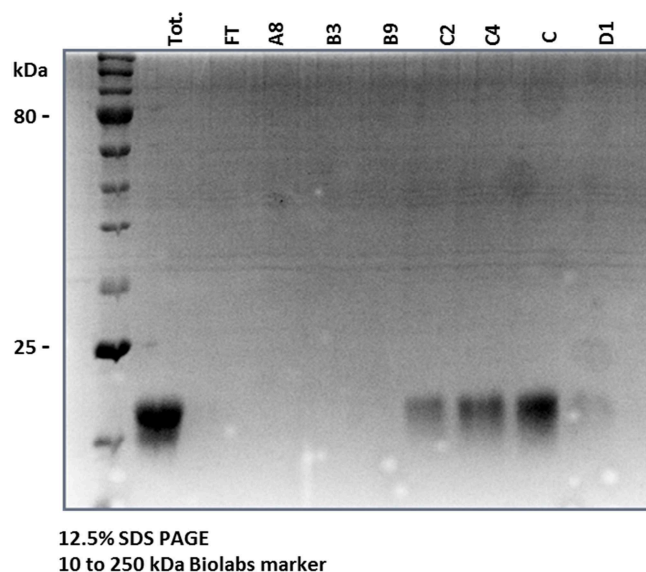


Figure 3.17: SDS PAGE analysis of fraction A8 to D1 from the Heparin columns, showing that HsLARP4(111-287) Q126A, of about 20.5 kDa was found predominantly from fraction C1 to D1. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane Tot.: Sample prior to 2nd round of purification, lane FT: Flow-through of protein solution after passing through the double Heparin columns, lane A8: Fraction A8, lane B3: Fraction B3, lane B9: Fraction B9, lane C2: Fraction C2, lane C8: Fraction C8, lane C4: Fraction C4, lane D1: Fraction D1.

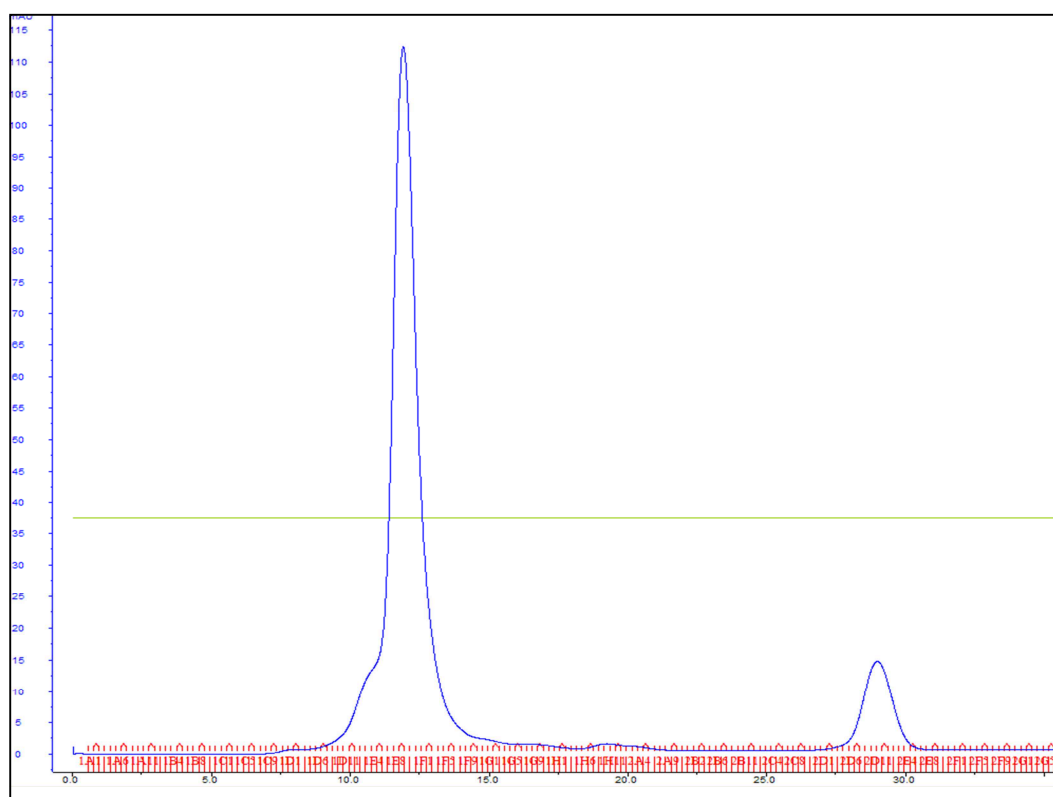


Figure 3.18: Elution profile HsLARP4(111-287) Q126A from the Size Exclusion column washed with NMR/ITC buffer . Protein elution was monitored by UV absorbance at 280nm . Distinct peak of UV absorbance was visible from fraction E5 to fraction F5. Blue curve: Ultraviolet Absorbance; Green curve: NMR/ITC buffer concentration.

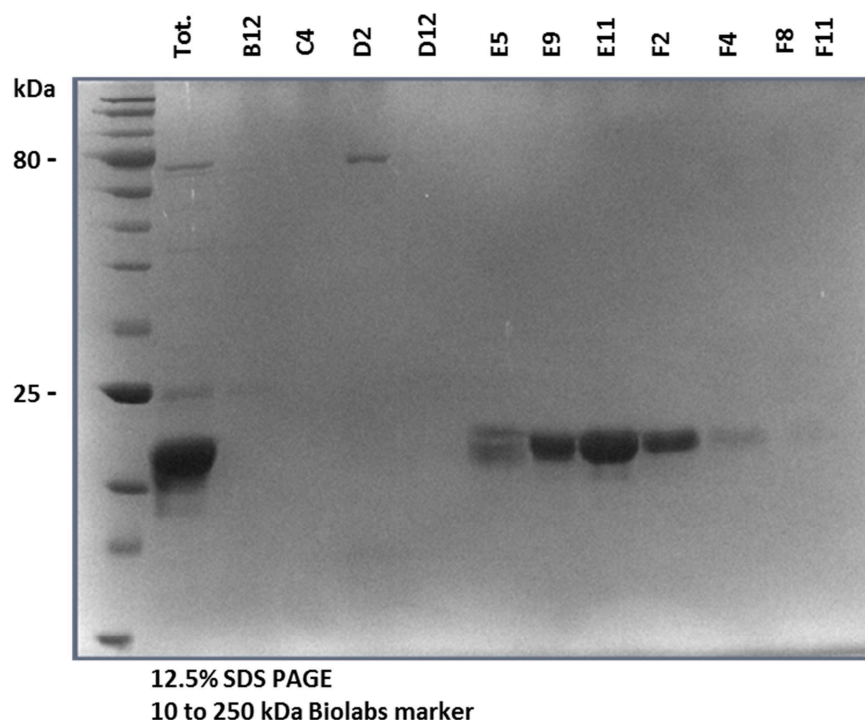


Figure 3.19: SDS PAGE analysis of fraction C4 to F11 from the size exclusion column, showing that HsLARP4(111-287) Q126A of about 20.5 kDa after removal of HisTag by TEV protease, was found predominantly from fraction E4 to F9. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane Tot.: Sample prior 3rd round of purification, lane B12: Fraction B12, lane C4: Fraction C4, lane D2: Fraction D2, lane D12: Fraction D12, lane E5: Fraction E5, lane E9: Fraction E9, lane E11: Fraction E11, lane F2: Fraction F2, lane F4: Fraction F4, lane F8: Fraction F8, lane F11: Fraction F11.

HsLARP4(111-287) D139A. Figure 3.20 shows the elution profile of HsLARP4(111-287) D139A from Nickel Column, with the peak occurring from fraction B2 to C2. SDS-PAGE analysis (Figure 3.21) confirmed the presence of the protein from fraction B2 to C2.

Manual purification through Nickel column after TEV cleavage removed the His-tag, TEV protease and contaminating protein. In Figure 3.22, the size of the protein reduced from about 23.1 kDa to 20.5 kDa after TEV cleavage.

Purification through Heparin columns removed nucleic acids contaminants. Figure 3.23 shows elution profile from Heparin Column, with the peak occurring

from fraction C1 to D5. SDS-PAGE analysis (Figure 3.24) confirmed the presence of the protein from fraction C3 to D3.

Finally, in order to eliminate last residues of contaminants, size exclusion column (SEC) purification was performed. Figure 3.25 shows the elution profile from SEC, with a major peak occurring from fraction E5 to F5. According to SDS-PAGE analysis (Figure 3.26), only the fractions from E5 to F5 were collected.

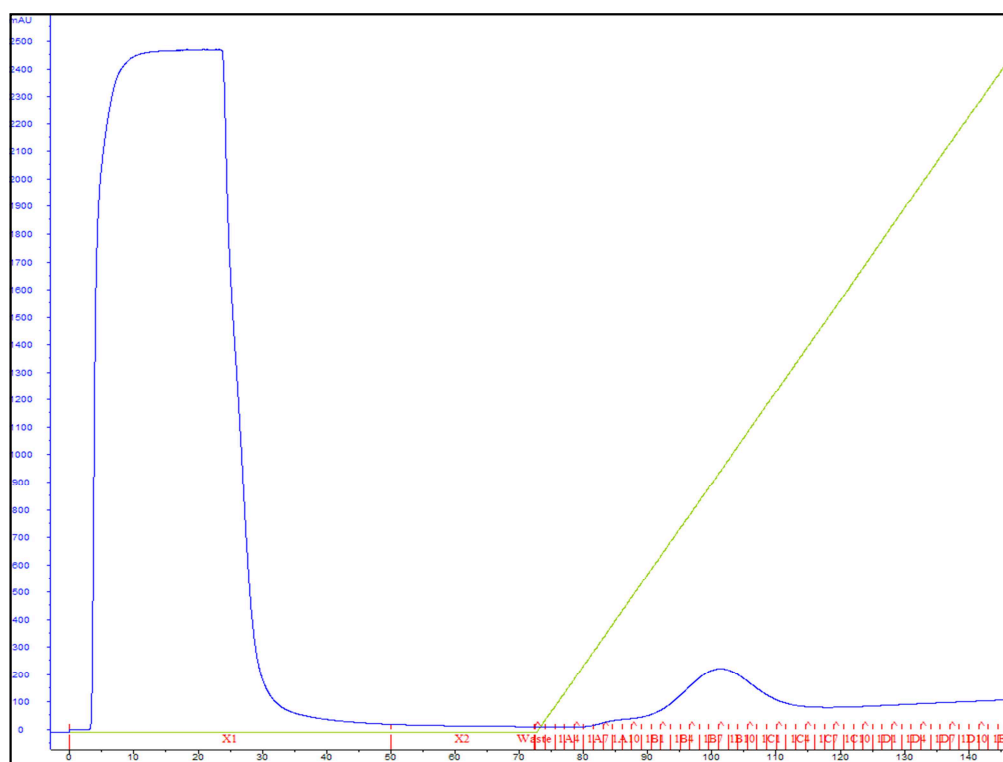


Figure 3.20: Elution profile HsLARP4(111-287) D139A from Nickel affinity column using a gradient of Nickel buffer B following a wash with Nickel buffer A. Protein elution was monitored by UV absorbance at 280nm. Distinct peak of UV absorbance was observed from fraction B2 to C2. Blue curve: Ultraviolet Absorbance; Green curve: Nickel Buffer B concentration.

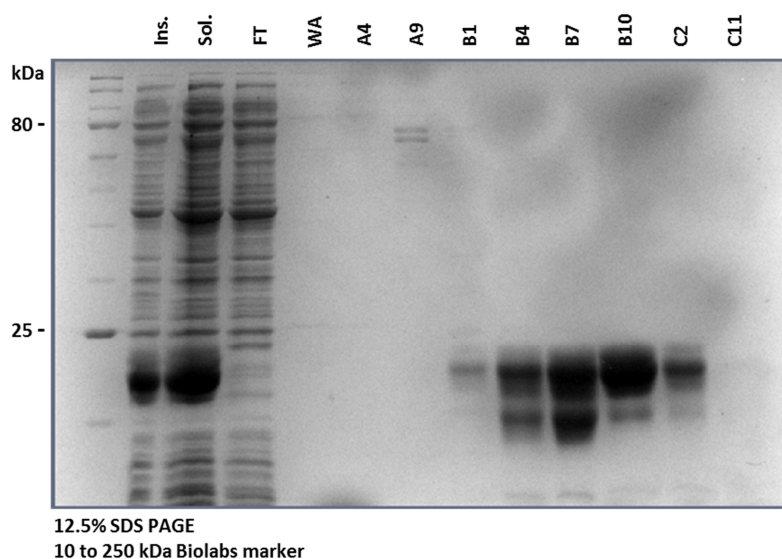


Figure 3.21: SDS PAGE analysis of fraction A4 to C11 from the Nickel column, showing that HsLARP4(111-287) D139A, of about 23.1 kDa, was found predominantly from fraction B2 to C2. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane Ins.: Insoluble cell fraction after sonication and centrifugation, lane Sol.: Soluble cell fraction after sonication and centrifugation lane FT: Flow-through from Nickel column, lane 5: Fraction collected after washing the column with Nickel buffer A, lane A4: Fraction A4, lane A9: Fraction A9, lane B1: Fraction B1, lane B4: Fraction B4, lane B7: Fraction B7, lane B10: Fraction B10, lane C2: Fraction c2, lane C11: Fraction C11.

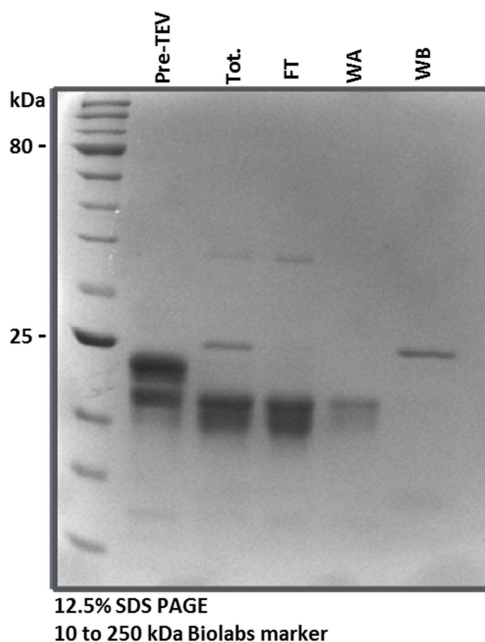


Figure 3.22: SDS-PAGE analysis of the protein solution after the 2nd round of purification through the Nickel affinity column after the removal of His-tag by TEV protease. Prior to TEV cleavage, the protein was about 23.1 kDa. After TEV cleavage, the molecular weight of the protein was reduced to about 20.5 kDa. lane 1: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane 2: protein solution prior to cleavage, lane 3: protein solution after cleavage, but prior to passing through the Nickel column, lane 4: protein solution after passing through the Nickel column and eluted out in the flow-through, lane WA: Elution of the contaminating protein by washing the column with Nickel Buffer A, lane WB: Elution of the contaminating protein by washing the column with Nickel Buffer B.

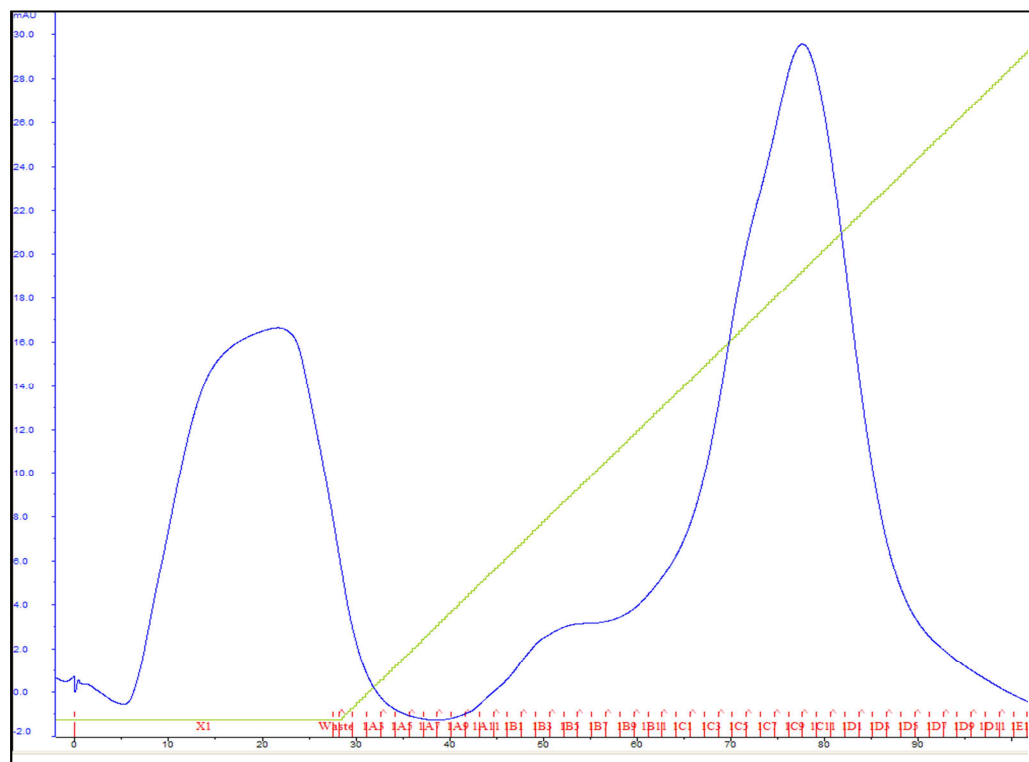


Figure 3.23: Elution profile HsLARP4(111-287) D139A from the Heparin columns washed by Heparin Buffer A, and eluted with a Heparin Buffer B gradient. Protein elution was monitored by UV absorbance at 280nm. Distinct peak of UV absorbance was formed from fraction C1 to fraction D5. Blue curve: Ultraviolet Absorbance; Green curve: Heparin Buffer B concentration.

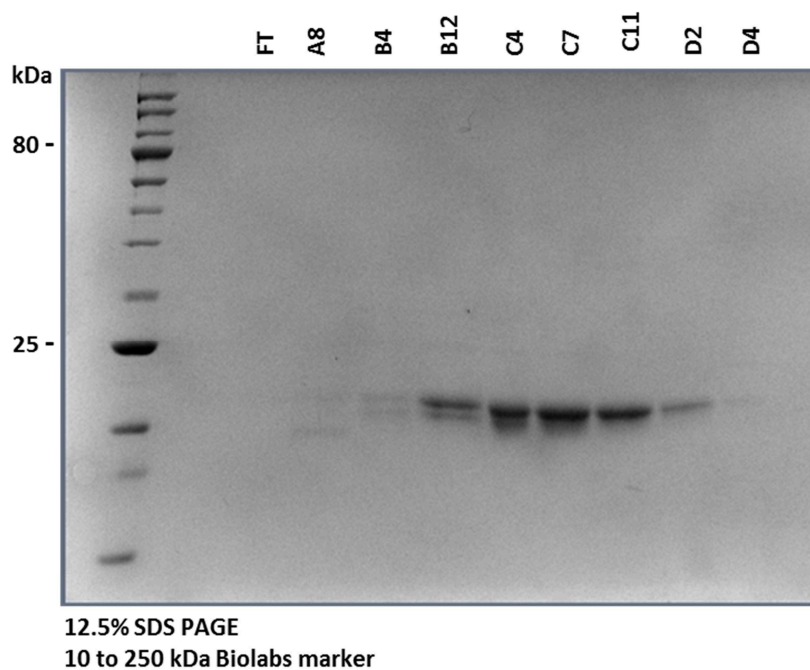


Figure 3.24: SDS PAGE analysis of fraction A8 to D4 from the Heparin columns, showing that HsLARP4(111-287) D139A, of about 20.5 kDa was found predominantly from fraction C3 to D3. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane FT: Flow-through of protein solution after passing through the double Heparin column, lane A8: Fraction A8, lane B4: Fraction B4, lane B12: Fraction B12, lane C4: Fraction C4, lane C7: Fraction C7, lane C11: Fraction C11, lane D2: Fraction D2, lane D4: Fraction D4.

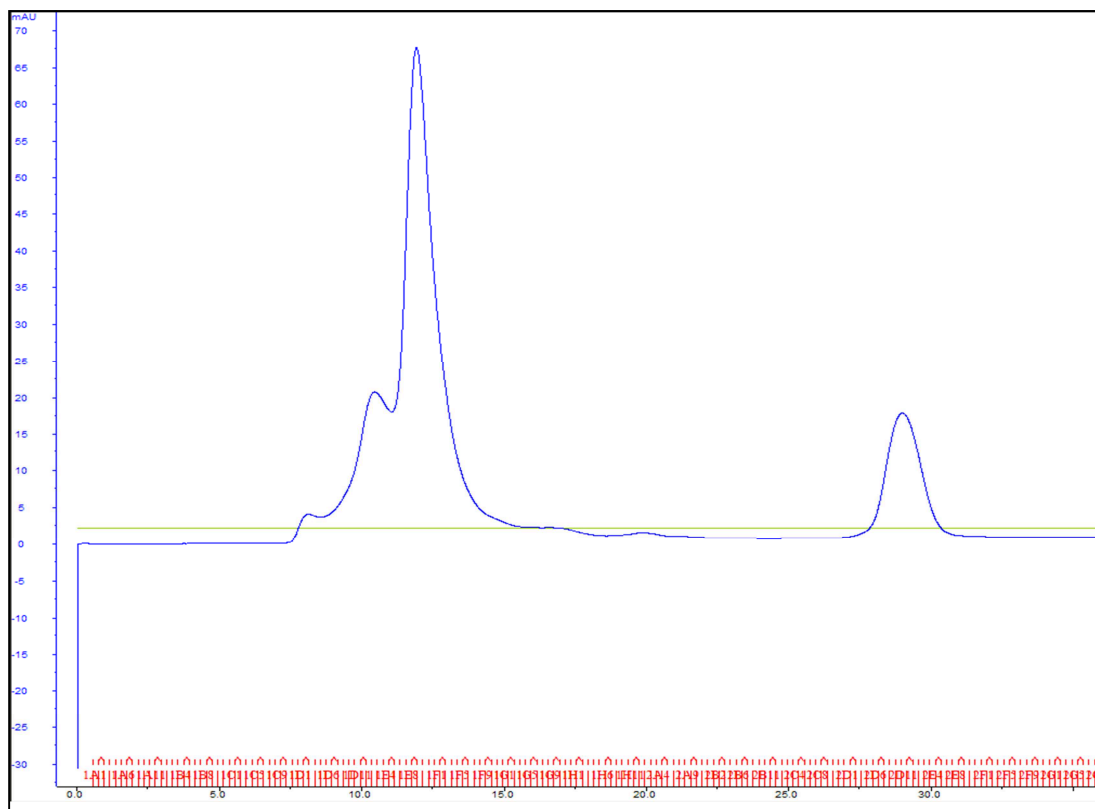
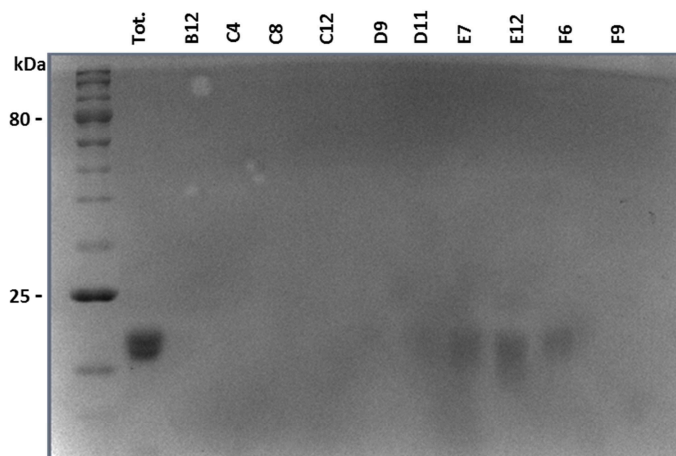


Figure 3.25: Elution profile HsLARP4(111-287) D139A from the Size Exclusion column washed with NMR/ITC buffer. Protein elution was monitored by UV absorbance at 280nm. A major peak of UV absorbance was visible from fraction E5 to fraction F5. Blue curve: Ultraviolet Absorbance; Green curve: NMR/ITC buffer concentration.



12.5% SDS PAGE
10 to 250 kDa Biolabs marker

Figure 3.26: SDS PAGE analysis of fraction C4 to F9 from the size exclusion column, showing that HsLARP4(111-287) D139A of about 20.5 kDa after removal of HisTag by TEV protease, was found predominantly from fraction E5 to F5. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane Tot.: Sample prior 3rd round of purification, lane B12: Fraction B12, lane C4: Fraction C4, lane C8: Fraction C8, lane C12: Fraction C12, lane D9: Fraction D9, lane D11: Fraction D11, lane E7: Fraction E7, lane E12: Fraction E12, lane F6: Fraction F6, lane F9: Fraction F9.

HsLARP4(111-287) M160A. Figure 3.27 shows the elution profile of *HsLARP4(111-287) M160A* from Nickel Column, with the peak occurring from fraction A5 to C2. SDS-PAGE analysis (Figure 3.28) confirmed the presence of the protein from fraction B2 to C1.

Manual purification through Nickel column after TEV cleavage removed the His-tag, TEV protease and contaminating protein. In Figure 3.29, the size of the protein reduced from about 23.1 kDa to 20.5 kDa after TEV cleavage.

Purification through Heparin columns removed nucleic acids contaminants. Figure 3.30 shows elution profile from Heparin Column, with the peak occurring from fraction B12 to D3. SDS-PAGE analysis (Figure 3.31) confirmed the presence of the protein from fraction C3 to D1.

Finally, in order to eliminate last residues of contaminants, size exclusion column (SEC) purification was performed. Figure 3.32 shows the elution profile from SEC, with a major peak occurring from fraction E6 to F6. According to SDS-PAGE analysis (Figure 3.33), only the fractions from E6 to F6 were collected.

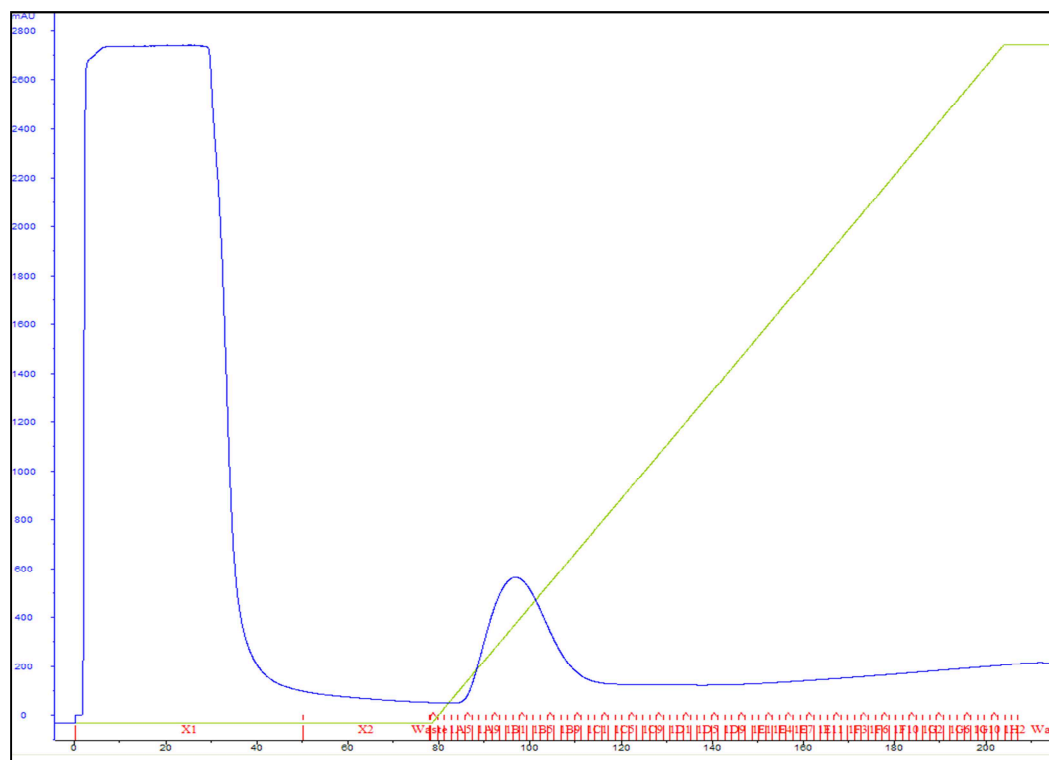


Figure 3.27: Elution profile HsLARP4(111-287) M160A from Nickel affinity column using a gradient of Nickel buffer B following a wash with Nickel buffer A. Protein elution was monitored by UV absorbance at 280nm. Distinct peak of UV absorbance was observed from fraction A5 to C2. Blue curve: Ultraviolet Absorbance; Green curve: Nickel Buffer B concentration.

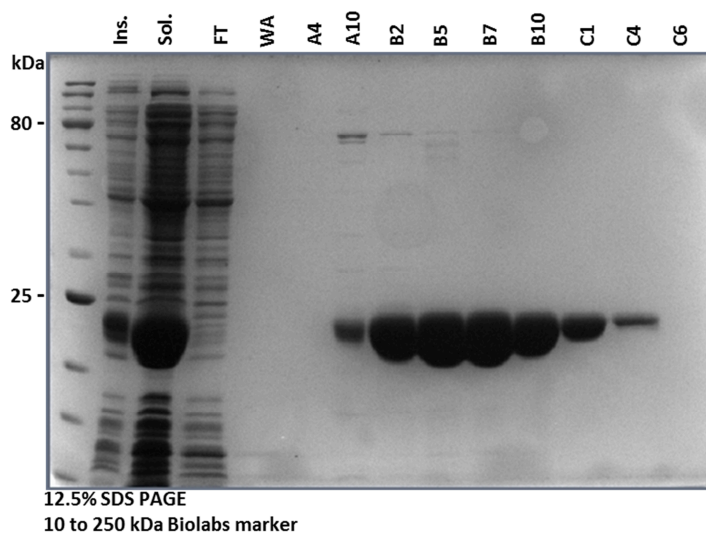


Figure 3.28: SDS PAGE analysis of fraction A4 to C6 from the Nickel column, showing that HsLARP4(111-287) M160, of about 23.1 kDa, was found predominantly from fraction B2 to C2. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane Ins.: Insoluble cell fraction after sonication and centrifugation, lane Sol.: Soluble cell fraction after sonication and centrifugation lane FT: Flow-through from Nickel column, lane 5: Fraction collected after washing the column with Nickel buffer A, lane A4: Fraction A4, lane A10: Fraction A10, lane B2: Fraction B2, lane B5: Fraction B5, lane B7: Fraction B7, lane B10: Fraction B10, lane C1: Fraction C1, lane C4: Fraction C4, lane C6: Fraction C6.

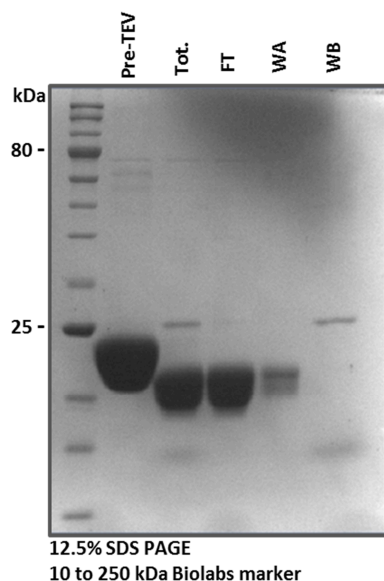


Figure 3.29: SDS-PAGE analysis of the protein solution after the 2nd round of purification through the Nickel affinity column after the removal of His-tag by TEV protease. Prior to TEV cleavage, the protein was about 23.1 kDa. After TEV cleavage, the molecular weight of the protein was reduced to about 20.5 kDa. lane 1: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane 2: protein solution prior to cleavage, lane 3: protein solution after cleavage, but prior to passing through the Nickel column, lane 4: protein solution after passing through the Nickel column and eluted out in the flow-through, lane WA: Elution of the contaminating protein by washing the column with Nickel Buffer A, lane WB: Elution of the contaminating protein by washing the column with Nickel Buffer B.

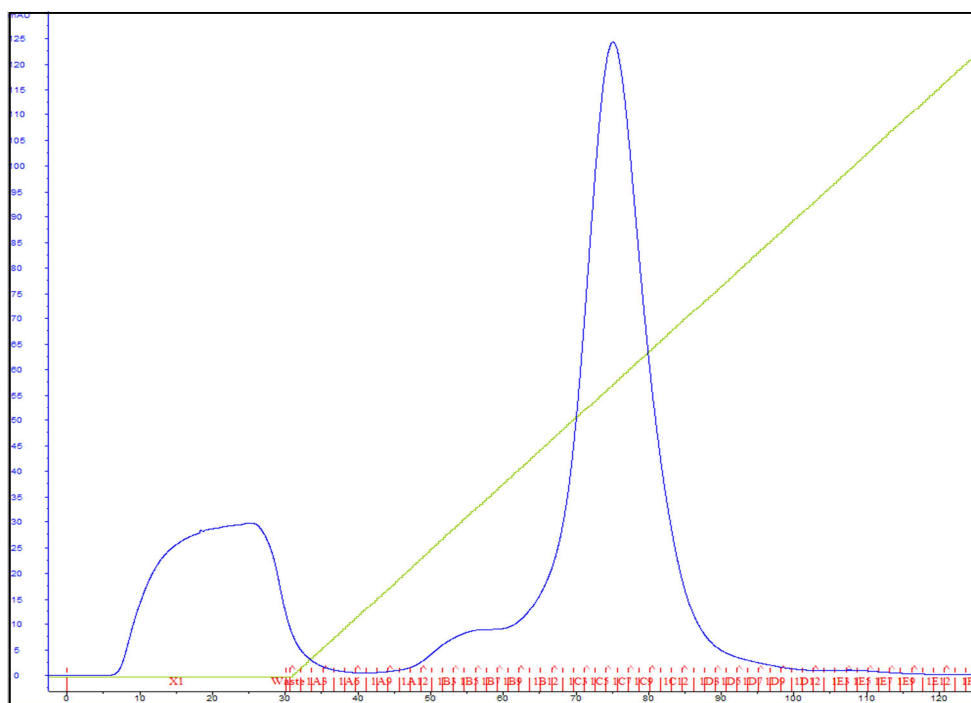


Figure 3.30: Elution profile HsLARP4(111-287) M160A from the Heparin columns washed with Heparin Buffer A, and eluted with a Heparin Buffer B gradient. Protein elution was monitored by UV absorbance at 280nm. Distinct peak of UV absorbance was formed from fraction B12 to fraction D3. Blue curve: Ultraviolet Absorbance; Green curve: Heparin Buffer B concentration.

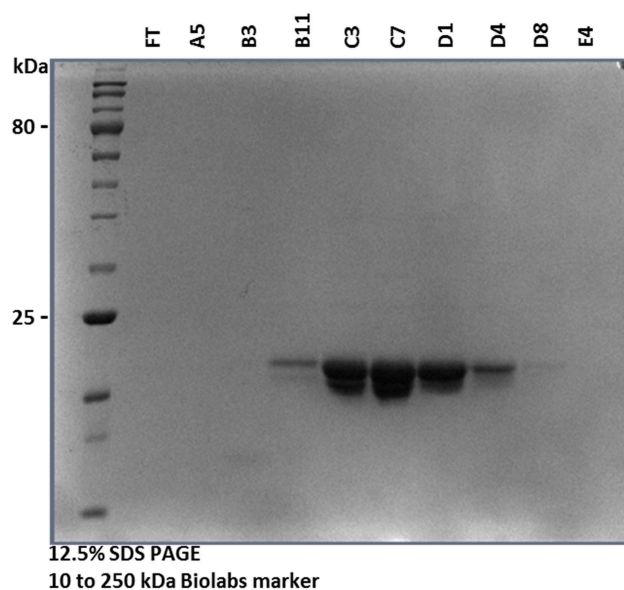


Figure 3.31: SDS PAGE analysis of fraction A5 to E4 from the Heparin columns, showing that HsLARP4(111-287) M160A, of about 20.5 kDa was found predominantly from fraction C3 to D1. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane FT: Flow-through of protein solution after passing through the double Heparin column, lane A5: Fraction A5, lane B3: Fraction B3, lane B11: Fraction B11, lane C3: Fraction C3, lane C7: Fraction C7, lane D1: Fraction D1, lane D4: Fraction D4, lane D8: Fraction D8, lane E4: Fraction E4.

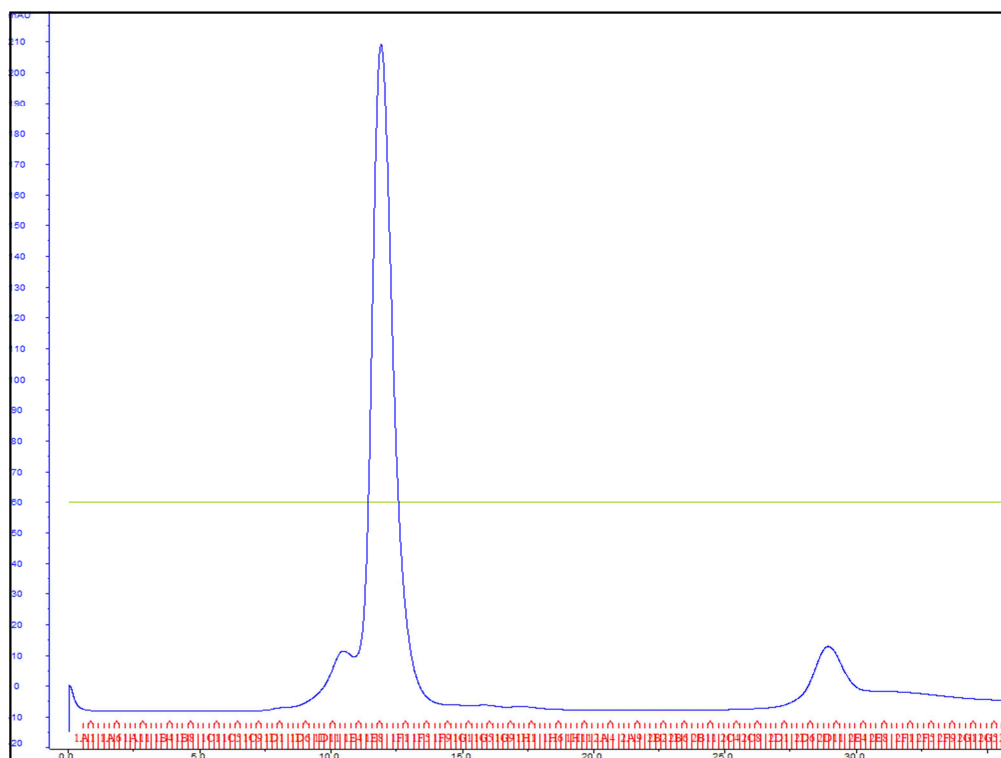


Figure 3.32: Elution profile HsLARP4(111-287) M160A from the Size Exclusion column washed with NMR/ITC buffer. Protein elution was monitored by UV absorbance at 280nm. A major peak of UV absorbance was visible from fraction E6 to fraction F6. Blue curve: Ultraviolet Absorbance; Green curve: NMR/ITC buffer concentration.

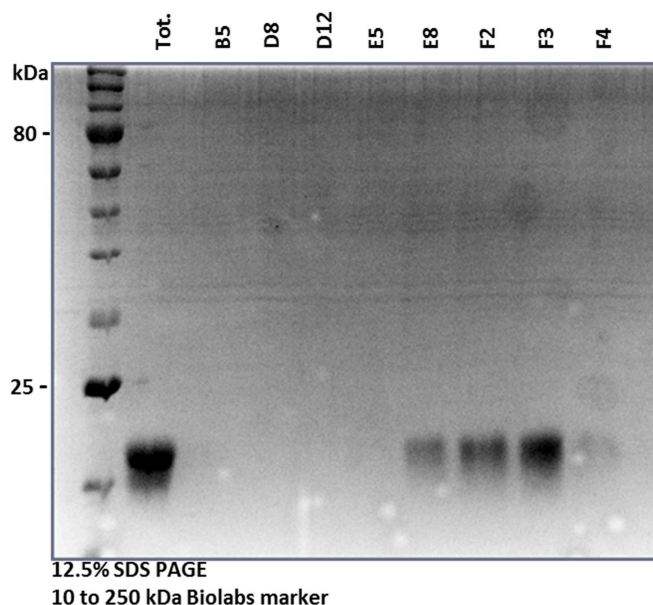


Figure 3.33: SDS PAGE analysis of fraction B5 to F4 from the size exclusion column, showing that HsLARP4(111-287) M160A of about 20.5 kDa after removal of HisTag by TEV protease, was found predominantly from fraction E6 to F6. First lane: Protein Ladder 10-250 kDa (New England BioLabs, UK), lane Tot.: Sample prior 3rd round of purification, lane B5: Fraction B5, lane D8: Fraction D8, lane D12: Fraction D12, lane E5: Fraction E5, lane E8: Fraction E8, lane F2: Fraction F2, lane F3: Fraction F3, lane F4: Fraction F4.

1D ^1H Nuclear Magnetic Resonance of HsLARP4(111-287) & the mutants.

1D ^1H -NMR was performed on HsLARP4(111-287), HsLARP4(111-287) Q126A, HsLARP4(111-287) D139A and HsLARP4(111-287) M160A to ascertain whether the purified proteins were folded properly.

A simple one-dimensional proton experiment, the most basic spectrum in NMR spectroscopy, already contains a great amount of information. An unfolded protein shows a small dispersion of the amide backbone chemical shift¹. Particularly, the appearance of intensities at chemical shifts near 8.3 ppm is an excellent indicator for a disordered protein, as this is a region of backbone amides in random-coil configuration. On the other hand signal dispersion beyond 8.5 ppm (8.5–11 ppm) proves a protein to be folded². Because of the different chemical

environment, and thus, the varying shielding effect, the resonances of the single protons will be distributed over a wide range of frequencies

In the NMR spectra of Figure 3.34, the presence of good chemical shift dispersion indicates that HsLARP4(111-287) and the three mutants assumed a folded structure in solution. Furthermore the 3 spectra appeared similar to each other, with minor differences in the chemical shift of some residues. This indicates that all the proteins could be folded in the same way, with minor differences in the local conformation.

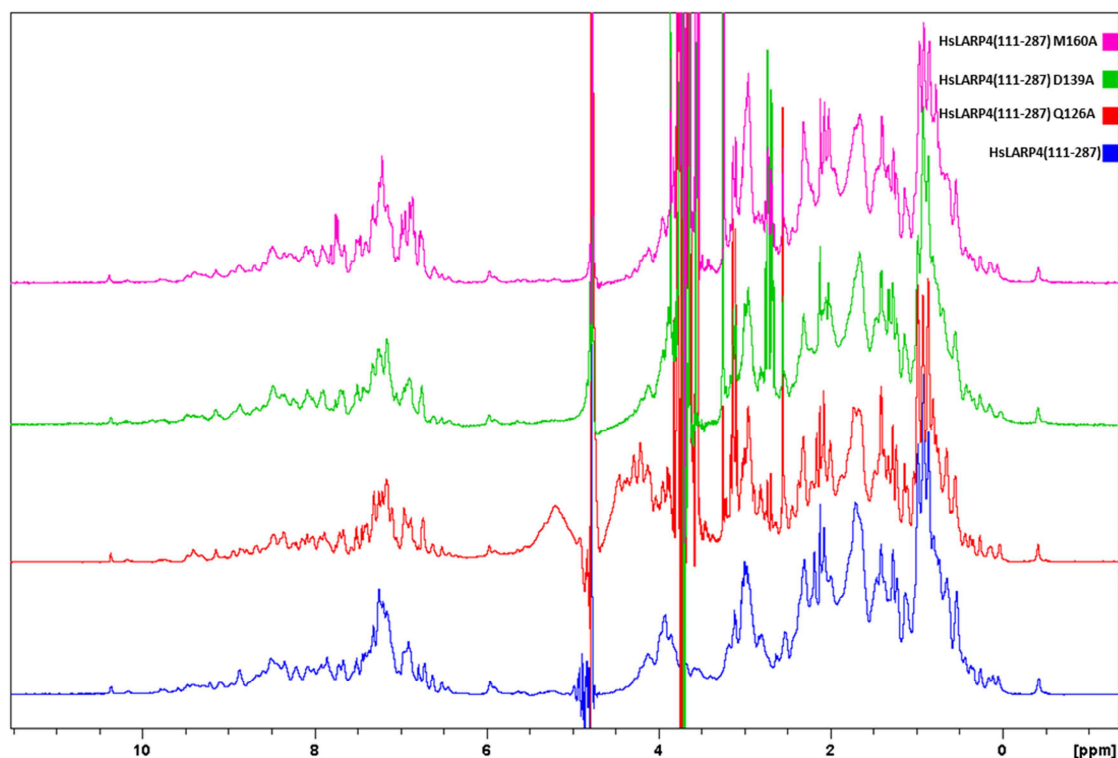


Figure 3.34: NMR spectra of HsLARP4(111-287) wild-type (Blue), mutant Q126A (Red), mutant D139A (Green) and mutant M160A (Pink).

Study of interaction between HsLARP4(111-287) protein and polyA-15 RNA using Isothermal Titration Calorimetry (ITC). ITC experiments were conducted to study the interaction of HsLARP4(111-287) and mutants HsLARP4(111-287) Q126A,

HsLARP4(111-287) D139A and HsLARP4(111-287) M160A with poly-A15 RNA. In the ITC experiment of HsLARP4(111-303) wild-type and poly-A15 RNA, binding occurs as one event centered on a molar ratio of 1 (Figure 2.4 of section 2.2). Further analysis revealed that LARP4(111-303) interacts with A(15) with a K_d of 714 nM, with entropically driven binding (Table 2.1 of section 2.2). When the protein in the cell was replaced with HsLARP4(111-287) Q126A, HsLARP4(111-287) D139A and HsLARP4(111-287) M160A, respectively, the interaction with poly-A15 RNA dramatically changed. Particularly, for all the three mutants the ITC experiments showed the absence of a binding curve, indicating that the proteins were no longer able to interact with the RNA (Figures 3.35, 3.36).

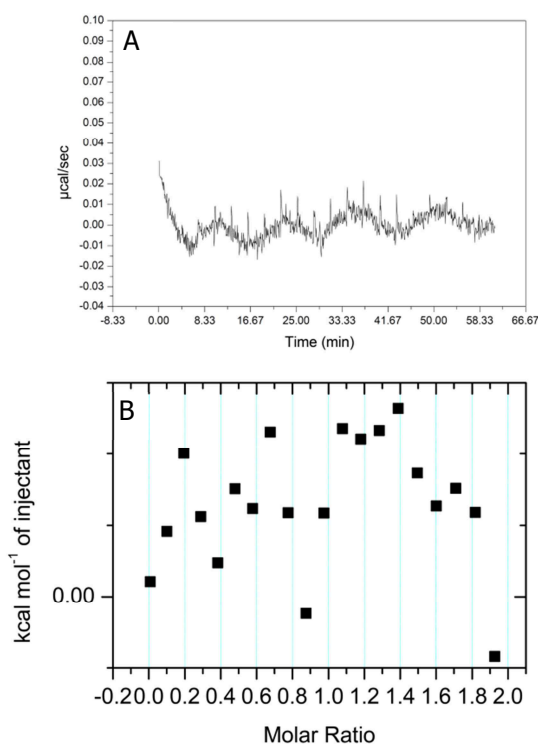


Figure 3.35. (A) Plot of heat of interaction for each injection of poly-A15 into cell containing mutant HsLARP4(111-287) Q126A. Peaks below the baseline indicate exothermic reactions. Peaks above the baseline indicate endothermic reactions. (B) Binding curve of wild-type HsLARP4(111-287) Q126A to poly-A15 RNA, fitted using the MicroCal-Origin 7.0, which integrates the peaks formed by the heat of interaction in (A).

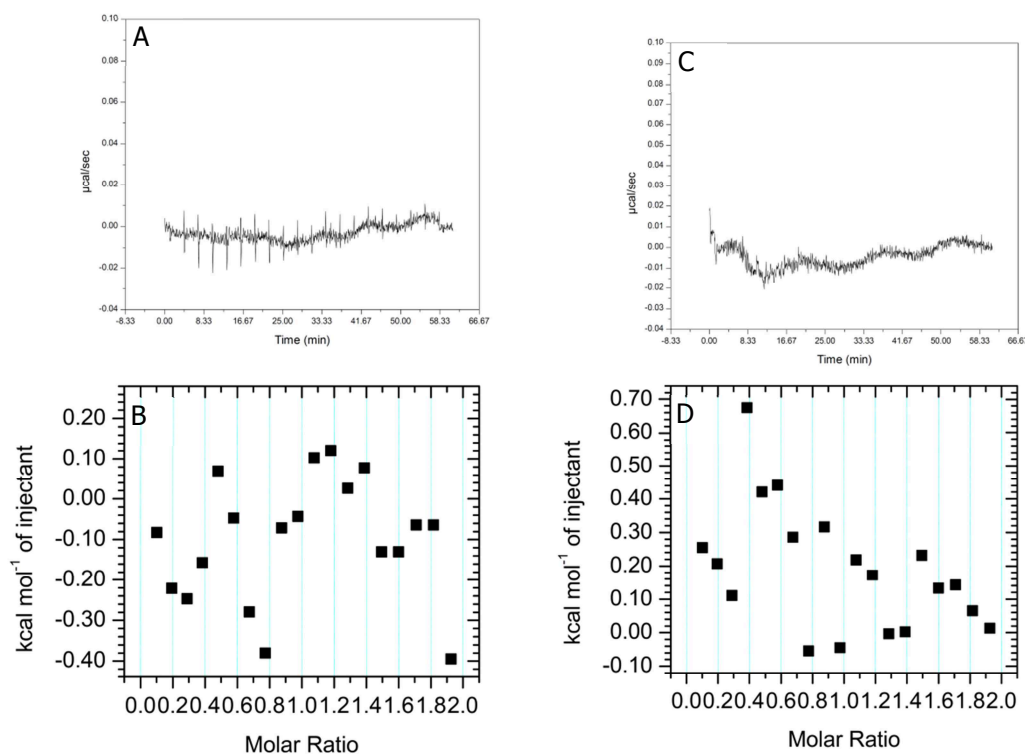


Figure 3.36. (A) Plot of heat of interaction for each injection of poly-A15 into cell containing mutant HsLARP4(111-287) D139A. Peaks below the baseline indicate exothermic reactions. Peaks above the baseline indicate endothermic reactions. (B) Binding curve of wild-type HsLARP4(111-287) D139A to poly-A15 RNA, fitted using the MicroCal-Origin 7.0, which integrates the peaks formed by the heat of interaction in (A). (C) Plot of heat of interaction for each injection of poly-A15 into cell containing mutant HsLARP4(111-287) M160A. (D) Binding curve of wild-type HsLARP4(111-287) M160A to poly-A15 RNA, fitted using the MicroCal-Origin 7.0, which integrates the peaks formed by the heat of interaction in (C).

3.3 Conclusions

In summary, the cloning and large scale production of HsLARP4(111-287)Q126A, HsLARP4(111-287)D139A and HsLARP4(111-287)M160 were successful performed as shown in section 3.2. 1D ^1H -NMR revealed that the three mutants assumed a folding. finally ITC experiments showed that of HsLARP4(111-287)Q126A, HsLARP4(111-287)D139A and HsLARP4(111-287)M160 were not able to interact with polyA-15 RNA. These results suggest that the mutation from residues Q126, D139 and M160 of HsLARP4(111-287) to Alanine interfered with the mechanism of

interaction with polyA-15 RNA, hence, resulting in the loss of the binding affinity. Therefore, these residues may probably play a key role for the protein activity.

3.4 Materials and Methods

Cloning of plasmid encoding for HsLARP4 mutants. cDNA encoding for HsLARP4 cDNA nucleotide 111 to 287 (HsLARP4(111-287)), Histidine-Tag (His-Tag) and Tobacco Etch Virus (TEV) cleavage site, were inserted into Novagen pETDuet-1 plasmid (Merck, Germany), and transformed into KRX cells. The plasmids were extracted using PureYield™ Miniprep (Promega, USA) and used as template for point-mutation. The three selected residues were mutated to Alanine through the Polymerase Chain Reaction (PCR) using primers in Figure 3.37. PCR was performed using the mixture composition reported in Table 3.1 and parameters reported in Table 3.2. After PCR, the templates were digested using 1 µL DpnI (New England BioLabs, UK).

HsLARP4 cDNA	122-tgtctgaagaa ca attagaattctgtttttcagagaaaatttgtcaaag gat ctttacttgata-143	155-acagttgccaac atg gaagaataaaaa-164
Q126A Forward Primer	5'tgtctgaagaa gc attagaattctgt 3'	
Q126A Reverse Primer	3'acagacttcttt cgt aattcttaagaca 5'	
D139A Forward Primer		5'aatttgtcaaag gct ctttacttgata 3'
D139A Reverse Primer		3'ttaacagttt cga gaatgaactat 5'
M160A Forward Primer		5'acagttgccaac gcc gaagaataaaaa 3'
M160A Reverse Primer		3'tgtcaacggtt cgg cttctttatattt 5'

Figure 3.37. Primers used for point mutation to generate the plasmids encoding for desired HsLARP4(111-287) mutants. Primers were designed to mutate Q126, D139 and M160 residues of HsLARP4(111-287) wild-type to Alanine through modifying the nucleotides of the corresponding codons.

Components	Volume (μ L)
Template (100ng/ μ l) – Plasmids encoding for Wild-type HsLARP4(111-287)	1
Forward primer (10 μ M) (Sigma-Aldrich, USA)	1
Reverse primer (10 μ M) (Sigma-Aldrich, USA)	1
dNTPs mix (25mM) (MP Biomedicals, United Kingdom)	0.4
10x cloned PFU reaction buffer (Agilent Technologies, USA)	5
Hot-start PFU turbo DNA polymerase (Agilent Technologies, USA)	1
Nuclease-free water	40.6

Table 3.1: Polymerase Chain Reaction Mixture Composition for Point Mutation

Programme	Temperature (oC)	Duration	Cycle
Polymerase activation	95	2 minutes	1
Denaturation	95	30 seconds	30
Annealing	55	30 seconds	
Elongation	72	6 minutes	
Ending	72	10 minutes	1

Table 3.2: Polymerase Chain Reaction parameters for Point Mutation to generate plasmids encoding for HsLARP4 mutants

Expression test of HsLARP4 proteins and the mutants. Plasmids reported in Figure 3.37 were transformed into *E.coli* KRX and Rosetta2 cells. Isolated colony was inoculated into 5 ml Luria-Bertani broth (Table 3.3) with 10 µg/ml ampicillin for KRX cells and with 10 µg/ml ampicillin and 34 µg/ml chloramphenicol for Rosetta2 cells, and incubated until broth became turbid. 500µL culture was used as “Pre-induction sample”. The rest were induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Calbiochem, USA) and 0.1% L-Rhamnose (Sigma-Aldrich, USA) for KRX cells and 1 mM IPTG for Rosetta2 cells, and incubated at 37 °C for three hours or 18 °C overnight. 500 µL culture was used as “Post-induction sample”. The rest were pelleted at 13200 rpm, re-suspended in 100 µL Nickel-buffer-A (Table 3.4) and sonicated for two seconds at amplitude of 12 microns for five times, with 20 seconds rest on ice between each sonication. The suspension was centrifuged at 13200 rpm for five minutes. 100 µL supernatant was used as “Soluble fraction sample”. The pellet was re-suspended in 60 µL 2% SDS solution and used as “Insoluble fraction sample”. The samples were analyzed using Sodium Dodecyl Sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Components	Amount / L
Tryptone Microbial Media (Fisher Chemical, UK)	10 g
Yeast Extract (Oxoid, UK)	5 g
Sodium Chloride (Fisher Scientific, USA)	10 g

Table 3.3: Composition of Luria Bertani (LB) Broth

Large scale production of HsLARP4(111-287) and the mutants. 1 mL 1 mM IPTG was added to 1 litre of Rosetta2 cells with an OD600 of 0.6 , and incubated at 18 °C overnight. It was centrifuged at 5000 rpm for 20 minutes, with the supernatant discarded, and the pellet resuspended in 30 ml Nickel buffer A (Table 3.4), 700 µL phenylmethanesulfonylfluoride (Sigma-Aldrich, USA) and 0.1 g lysozyme (Sigma-Aldrich, USA). Pulse-sonication of five seconds at amplitude of 10-14 microns, and six seconds rest on ice was done for five minutes, and repeated again. After centrifuging at 18,000 rpm for 45 minutes, the supernatant was filtered with a 0.5 µm filter and injected into a Nickel column on the Akta system (GE Healthcare, UK). Following Nickel buffer A wash with 10 column volumes (CV), the protein was eluted with a linear gradient from 0 up to 80% of buffer B (Table 3.5) in 25 CV. Eluent was treated with 1.5 ml 0.05 mg/ml TEV protease, followed by 4 °C overnight buffer exchange with Dialysis-buffer in Spectra/Por® membrane (MWCO: 6-8000) (Spectrum Laboratories, US). Nickel column chromatography was manually repeated, followed by buffer exchange using Dialysis-buffer (Table 3.6). Heparin column chromatography was performed using Heparin buffers A and B (Table 3.7) eluting LARP4 by a linear gradient of buffer B gradient from 0 up to 100% of buffer B (Table 3.8) in 30 CV After 4 °C overnight buffer exchange with NMR/ITC buffer (Table 3.9), LARP4 proteins were purified using Size Exclusion Column (SEC) using the NMR/ITC buffer isocratically. Fractions corresponding to the peak of the protein were collected reaching a volume around 2 ml. Protein concentration was done using Vivaspin20 column (Sartorius Stedim Biotech, Germany). Nuclear Magnetic Resonance (NMR) spectroscopy was used to assess the correct fold of the purified proteins.

Components	Amount / L
Tris-base (Fisher Scientific, UK)	6.06 g
Sodium Chloride (Fisher Scientific, USA)	8.77 g
Imidazole (Acros Organic, Belgium)	0.68 g
Glycerol (Fisher Scientific, USA)	50 g
Milli-Q water (Millipore, USA)	Top up to 1L

Table 3.4: Composition of Nickel Buffer A (pH: 8)

Components	Amount / L
Tris-base (Fisher Scientific, UK)	6.06 g
Sodium Chloride (Fisher Scientific, USA)	8.77 g
Imidazole (Acros Organic, Belgium)	34.04 g
Glycerol (Fisher Scientific, USA)	50 g
Milli-Q water (Millipore, USA)	Top up to 1L

Table 3.5: Composition of Nickel Buffer B (pH: 8)

Components	Amount / 4L
Tris-base (Fisher Scientific, UK)	9.69 g
Potassium Chloride (Fisher Scientific, UK)	29 g
Dithiothreitol (Thermo Scientific, USA)	0.6 g
EDTA (Fisher Scientific, UK)	0.3 g
Milli-Q water (Millipore, USA)	Top up to 1L

Table 3.6: Composition of Dialysis Buffer (pH: 8)

Components	Amount / L
Tris-base (Fisher Scientific, UK)	6.06 g
Glycerol (Fisher Scientific, USA)	100 g
Milli-Q water (Millipore, USA)	Top up to 1L

Table 3.7: Composition of Heparin Buffer A (pH: 7.25)

Components	Amount / L
Tris-base (Fisher Scientific, UK)	6.06 g
Glycerol (Fisher Scientific, USA)	100 g
Potassium Chloride (Fisher Scientific, UK)	149.08 g
Milli-Q water (Millipore, USA)	Top up to 1L

Table 3.8: Composition of Heparin Buffer B (pH: 7.25)

Components	Amount/ 4L
Tris-base (Fisher Scientific, UK)	9.71 g
Potassium Chloride (Fisher Scientific, UK)	29.82 g
Dithiothreitol (Thermo Scientific, USA)	0.3 g
EDTA (Fisher Scientific, UK)	0.62 g
Milli-Q water (Millipore, USA)	Top up to 1L

Table 3.9 :Composition of NMR/ITC Buffer (pH: 7.25)

Study of interaction between HsLARP4(111-287) mutants and poly-A15

RNA using Isothermal Titration Calorimetry (ITC). Using the iTC200 system (GE Healthcare, UK), 20 computer-controlled injections of 2 μ L RNA (10-20 μ M) were injected into the cell containing the protein (80-160 μ M), with an interval of three minutes between each injection at 298 K. Both RNA and protein were dissolved in NMR/ITC buffer (Table 3.9). Heat data obtained from the titration were fitted to a theoretical binding curve using the software, MicroCalOrigin7.0.

References

- [1] Wuthrich K., *NMR of Proteins and Nucleic Acids* (New York: Wiley) **1986**.
- [2] Rehm T., Huber R., Holak T. A. *Structure*. **2002**;10, 1613-8.